ANALYSIS OF FEATURES FOR SYNTHETIC APERTURE RADAR TARGET CLASSIFICATION

THESIS

Aaron K. McCauley, First Lieutenant, USAF

AFIT-ENG-MS-15-M-032

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT-ENG-MS-15-M-032

ANALYSIS OF FEATURES FOR SYNTHETIC APERTURE RADAR TARGET CLASSIFICATION

THESIS

Presented to the Faculty

Department of Electrical and Computer Engineering

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

in Partial Fulfillment of the Requirements for the

Degree of Master of Science in Electrical Engineering

Aaron K. McCauley, B.S.E.E.

First Lieutenant, USAF

March 2015

AFIT-ENG-MS-15-M-032

ANALYSIS OF FEATURES FOR SYNTHETIC APERTURE RADAR TARGET CLASSIFICATION

THESIS

Aaron K. McCauley, B.S.E.E.
First Lieutenant, USAF

Committee Membership:

Dr. Julie Jackson
Chair

Major Brian Woolley, PhD
Member

Major Michael Seal, PhD
Member

AFIT-ENG-MS-15-M-032

Abstract

Considering two classes of vehicles, we aim to identify the physical elements of the vehicles with the most impact on identifying the class of the vehicle in synthetic aperture radar (SAR) images. We classify vehicles using features, from polarimetric SAR images, corresponding to the structure of physical elements. We demonstrate a method which determines the most impactful features to classification by applying subset selection on the features. Determination of the most impactful elements of the vehicles is beneficial to the development of low observables, target models, and automatic target recognition (ATR) algorithms.

We show how previous work with features from individual pixels is applied to a greater number of target states. At a greater number of target states, the previous work has poor classification performance. Additionally, the nature of the features from pixels limits the identification of the most impactful elements of vehicles. We apply concepts from optical sensing to reduce the limitation on identification of physical elements.

We draw from optical sensing feature extraction with the use of Histogram of Oriented Gradients (HOG). From the cells of HOG, we form features from frequency and polarization attributes of SAR images. Using a subset set of features, we achieve a classification performance of 96.10 percent correct classification. Using the features from HOG and the cells, we identify the features with the most impact.

Using backward selection, a process for subset selection, we identify the features with the most impact to classification. The execution of backward selection removes the features which induce the most error in classification. We report features extracted from polarization attributes of SAR images have the most positive impact on classification performance.

To my faithful and supportive wife, thank you for your support and continual reminders to rest in our Lord and Savior now and into the future.
To my parents who raised me in the Word of the Lord and taught, by example, proper work ethic.

Acknowledgments

I would like to thank my advisor, Dr. Julie Jackson, for her support and guidance throughout the culmination of this thesis. Without her technical expertise and experience, the work in this thesis would not be possible.

I would also like to thank my committee, Major Michael Seal and Major Brian Woolley, for the many hours of dedicated input throughout the thesis process. Their support enabled the pursuit of machine learning techniques and the High Performance Computing resources.

This work was supported in part by a grant of computer time from the DoD High Performance Computing Modernization Program at the AFRL DSRC at Wright Patterson.

This thesis research was sponsored by the Air Force Office of Scientific Research under lab task LRIR12RY19COR.

Aaron K. McCauley

Table of Contents

Page

List of Figures

List of Tables

List of Common Terms

Term	Definition
Attribute	Quantifiable information extractable from the target of interest. Attribute may be extracted from any source of measurable information. The attributes extracted from a source are a quantifiable measure of the source.
Element	A physical piece of a target, definable separate from the target.
Feature	A processed form of an attribute. May either be identical to an attribute or a quantifiable combination of multiple attributes.
Feature Vector	A combination of features extracted from the target of interest where the order of the features in the vector matters.

List of Acronyms

Acronym	Definition
AFRL	Air Force Research Laboratory
ATR	automatic target recognition
CFAR	constant false alarm rate
CV	civilian vehicles
HOG	histograms of oriented gradients
HPC	high performance computing
LDA	linear discriminant analysis
PEC	perfect electric conductor
RVM	relevance vector machine
SAR	synthetic aperture radar
SPLIT	spectrum parted linked image test
SVM	support vector machine

ANALYSIS OF FEATURES FOR SYNTHETIC APERTURE RADAR TARGET CLASSIFICATION

I. Introduction

Considering two classes of vehicles in Figure 1.1, we aim to identify the physical elements of the vehicles with the most impact on identifying the class from synthetic aperture radar (SAR) images. We classify vehicles using features, extracted from polarimetric SAR images, corresponding to the structure of physical elements. We demonstrate a method which determines the most impactful features to classification by applying a subset selection on the features. Determination of the most impactful elements of the vehicles is beneficial to the development of low observables, target models, and automatic target recognition (ATR) algorithms.

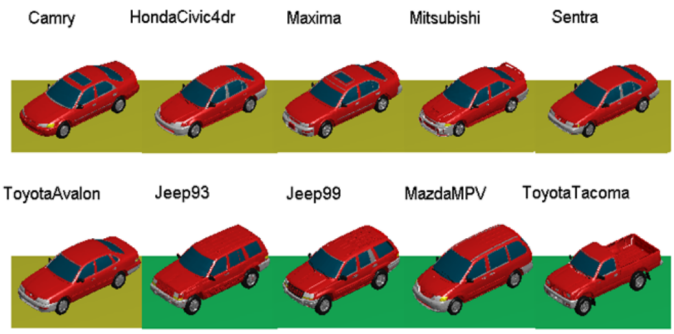

Figure 1.1: CV Domes Vehicles [1].

Various processes are used to identify targets within SAR images. Some processes chip the SAR images and then correlate the chips with a dictionary of chips to identify the class [2]. Other processes use Principle Component Analysis to define the SAR images and

a nearest neighbor classifier to identify the class [3, 4]. The diameter, inertia, percent bright constant false alarm rate (CFAR), and fractal dimension of the target in the image have been used and compared to training data to determine the class [5]. Another process fingerprints SAR images, utilizing machine learning to identify the class [6]. The previously stated research all assess the information from the entire SAR image to identify a target class. Work performed by Flynn [7] used information from a single pixel in the SAR image and machine learning to identify the class.

This thesis explores the pixel based features used by Flynn [7] and full image classification. Flynn's pixel based features tie information back to physical elements of the vehicles, but does not classify using the entire image. Full image classification utilizes the entire SAR image, but is limited in tying the features used to physical elements of the vehicle. We extend the work performed by Flynn [7] to extract features from the entire image. We also explore the impact of features on classification of the vehicle through the use of backward selection.

SAR ATR is a challenging problem. Inherent to the nature of SAR, the images suffer from low resolution compared to other imaging systems [8]. Also, the targets contain multiple states encompassing 360 degrees in azimuth and a change in elevation dependent upon the concept of operations. As such, classification of SAR images requires a high-dimensional feature space, which is computationally intensive. To evaluate classification performance, the Air Force Research Laboratory (AFRL) high performance computing (HPC) resources are used in the execution of this research [9].

Through subset selection, we show that classification performance is improved using features extracted from cells of an image. Subset selection also identifies features with the most impact on classification. Analysis of these features may expose the impact of physical elements of the vehicles on SAR ATR.

1.1 Thesis Organization

This thesis is organized into five chapters. Chapter II covers the tools, techniques, and algorithms used to support the analysis of the impact of features. Previous research by Flynn [7] is also reviewed in Chapter II. Chapter III covers the application of pixel and cell methods for feature extraction from SAR images. Flynn's research [7] is extended in Chapter III. Chapter IV covers the methodology behind backward selection and the evaluation of the impact of features. Chapter V completes this thesis with final conclusions and recommendations for future work.

II. Background

The goal of this chapter is to build and define the foundation of knowledge used in the development and execution of this research. Section 2.1 introduces the ATR process and focuses this thesis on feature formation. Section 2.2 introduces the nature of features, attributes, and feature vectors. Section 2.3 introduces the spectrum parted linked image test (SPLIT) algorithm for attribute extraction from SAR images. Section 2.4 introduces the classifiers we use in this thesis. Section 2.5 reviews how previous work by Flynn [7] formed features and used classifiers with SAR images.

2.1 ATR Overview

The ATR process is characterized by the process shown in Figure 2.1. Data is processed into a form where targets are detected. The detected targets are segmented from the data and features are extracted. Using the features, the target is classified. The classification of the target is used to impact system or mission parameters in real time [10]. In this thesis, we focus on the fourth stage of the ATR process (i.e. feature computation, selection, and classification).

Figure 2.1: Block Diagram of a Typical ATR System [10].

2.2 Feature Computation, Selection, and Classification

Features, formed from the attributes of SAR images, are applied in the ATR process to classify targets. Chapter III evaluates two ways to form features from the attributes in SAR

images. The full set of features used to classify targets is a feature vector. The relationship between attributes, features, and feature vectors is illustrated in Figure 2.2.

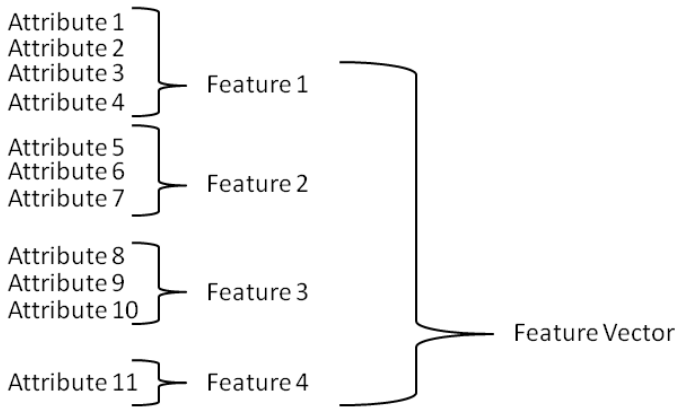

Figure 2.2: Relation of Attributes, Features, and Feature Vectors.

Figure 2.3 shows the generic process we use in this thesis to evaluate the classification performance of feature vectors. The first step extracts attributes from the data. The second step formulates the extracted attributes into features. Third, a feature vector is populated from a subset of all possible features. The fourth step is to train and test on the feature vector using a classifier to arrive at a metric of performance for the feature vector.

In Chapter III, we use SPLIT to extract the odd bounce polarization response attribute, k_o, from SAR images. Using the pixel method, we form a feature directly from the extracted k_o pixel value. Using the cell method, we form a feature from the mean of the extracted k_o pixel values over a spatial region of pixels called a cell. Other features are formed from different attributes. A combination of features forms a feature vector. The percent correct classification of targets is used as the metric of classification performance corresponding to the feature vector.

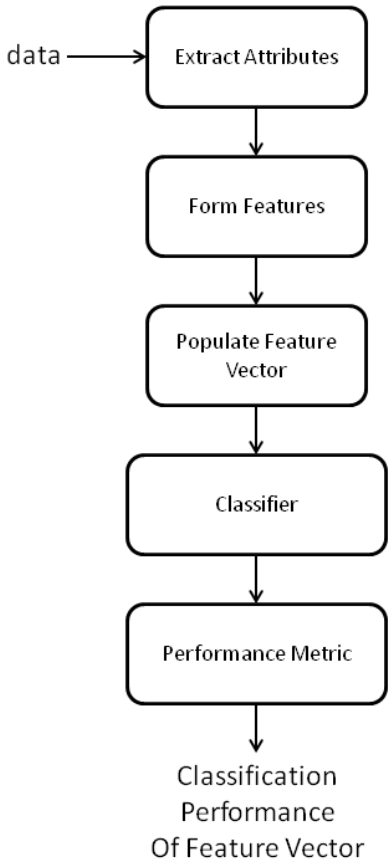

Figure 2.3: Block Diagram of Classification Process.

2.3 Attribute Extraction

In this thesis, we use the SPLIT algorithm developed by Fuller and Saville [11] for the extraction of attributes. SPLIT constructs a set of three 2-D sub-images for each polarization channel of the target in the x-y plane using a form of back-projection [11]. The sub-images are related to the frequency spectrum of the phase history. The phase history is filtered into overlapping frequency banks. The first frequency bank is the first half of the bandwidth, the second is the middle two-fourths of the bandwidth, and the third is the second half of the bandwidth.

Peaks that are a result of canonical scatterers are stable in location across sub-images of the back projection [12]. Attributes are extracted for the stable peak pixels in the sub-

images. As such, identification of the stable peaks in a target area is the first step in attribute extraction. SPLIT uses a watershed technique on the sub-images to identify peaks within each sub-image. The threshold of the watershed technique is variable within SPLIT. A low threshold rejects peaks with an amplitude more than 1dB below the maximum peak in the image. A medium threshold rejects peaks with an amplitude more than 10 dB below the maximum peak in the image. A high threshold rejects peaks with an amplitude more than 32 dB below the maximum peak. The frequency response attribute is extracted from the peak pixels. We use the high threshold in this thesis. Work with ATR and SPLIT reports classification performance is best using a high threshold [13].

SPLIT extracts α from each co-polarization channel, and $[k_o, k_e]$ from each of the sub-images [11]. The final α for the pixel is the weighted average of the attribute across the co-polarization channels. The weight of each α is the magnitude of the pixel amplitude related to the α. The final $[k_e, k_o]$ for the pixel is the weighted average of the attributes across the sub-images. Each attribute is described in detail in Subsections 2.3.1-2.3.3.

2.3.1 Amplitude Attribute.

The amplitude attributes of the pixels relate to the scatters of a target and are displayed as SAR images. SPLIT forms images from the horizontal polarization, P_{HH}, vertical polarization, P_{VV}, and cross polariztion, P_{HV}, channels using a form of backprojection [11]. The combination of the three images, $[P_{HH}, P_{VV}, P_{HV}]$, forms the final image, I, where the pixel amplitudes are the extracted amplitude attributes for the image. The three images are combined as [11]

$$I = |P_{HH}|^2 + |P_{VV}|^2 + |P_{HV}|^2. \tag{2.1}$$

2.3.2 Frequency Response Attribute.

The frequency response attribute is extracted from the change in pixel amplitude across the sub-images. The frequency response ties back to the curvature of the physical element of the target. Physical elements that are doubly curved, such as a sphere, have an

approximate α value of zero [11, 14]. Physical elements that are singly curved, such as a cylinder, have an approximate α value of one or negative one [11, 14]. Finally, target elements that are a corner reflector, such as a trihedral, have an approximate α value of two or negative two [11, 14]. The isotropic model for the frequency response is shown as [11, 15, 16]

$$\tilde{S}_f(f, A, \alpha) = A(jf)^{\frac{\alpha}{2}}, \tag{2.2}$$

where f is the frequency of the waveform, and A is a complex value related to the radar cross section of the point. The relationship between curvature and the α value is illustrated in Figure 2.4. SPLIT uses an iterative curve fitting algorithm to estimate α. The amplitude of the pixels of the three sub-images defines the curve α is estimated to fit. An iterative curve fitting method is used to minimize the residual between the estimated amplitude curve and measured amplitudes of the sub-images [11].

The α attribute for a pixel is fit to minimize the norm of the residual for the kth iteration expressed as $\left\| \frac{\boldsymbol{\sigma}}{\hat{\nu}_k} - \boldsymbol{f}(\acute{\alpha}_k) \right\|_2$ [11]. The normalization frequency vector, $\mathbf{f}(\alpha_k)$, is expressed as [11]

$$\mathbf{f}(\alpha_k) = \frac{[(f_{c1})^{\alpha_k+2}, (f_{c2})^{\alpha_k+2}, (f_{c3})^{\alpha_k+2}]^T}{(f_c)^{\alpha_k+2}}, \tag{2.3}$$

where f_{c1} is the center frequency of subimage 1, f_{c2} is the center frequency of subimage 2, f_{c3} is the center frequency of subimage 3, and f_c is the center frequency of the full bandwidth. The observation vector, $\boldsymbol{\sigma}$, is expressed as [11]

$$\boldsymbol{\sigma} = [|a_1|^2, |a_2|^2, |a_3|^2]^T, \tag{2.4}$$

where a_1 is the amplitude of the pixel in subimage 1, a_2 is the amplitude of the pixel in subimage 2, and a_3 is the amplitude of the pixel in subimage 3. The normalization factor, $\hat{\nu}_k$, is expressed as [11]

$$\hat{\nu}_k = \frac{\boldsymbol{\sigma}^T \boldsymbol{\sigma}}{\boldsymbol{\sigma}^T \boldsymbol{f}(\acute{\alpha}_k)}. \tag{2.5}$$

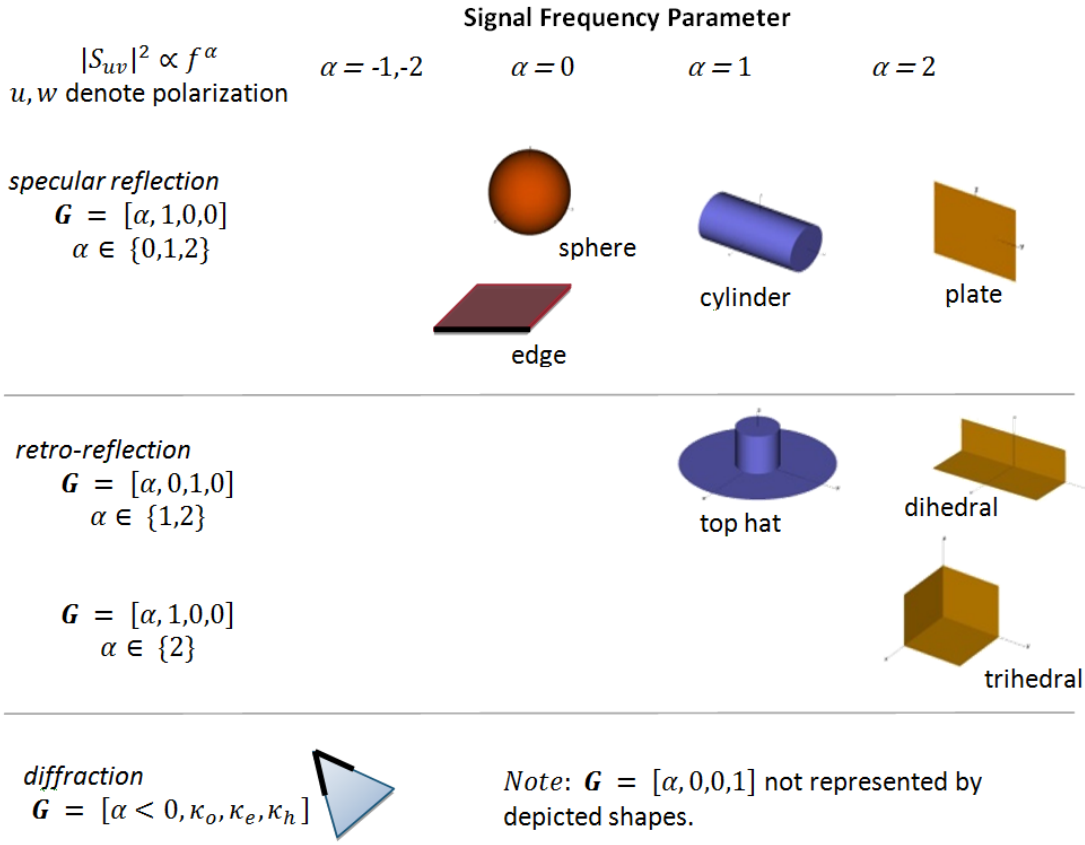

Figure 2.4: Relationship Between Curvature and Frequency Response (used with permission: Dr. Julie Jackson) [17].

The first iteration curve fitting method is initialized at [11]

$$\acute{\alpha}_1 = \frac{\log \frac{|a_1|^2}{|a_3|^2}}{\log \frac{f_{c1}}{f_{c3}}} - 2. \tag{2.6}$$

The frequency parameter, α_k, is adjusted by a scaled version of norm of the residual expressed as [11]

$$\delta_k = (0.95)^k \left\| \frac{\boldsymbol{\sigma}}{\hat{\boldsymbol{v}}_k} - \boldsymbol{f}(\acute{\alpha}_k) \right\|_2, \tag{2.7}$$

where α_{k+1} is expressed as

$$\acute{\alpha}_{k+1} = \begin{cases} \acute{\alpha}_k + \delta_k, & \left\| \frac{\boldsymbol{\sigma}}{\hat{\boldsymbol{v}}_k} - \boldsymbol{f}(\acute{\alpha}_k + \delta_k) \right\|_2 < \left\| \frac{\boldsymbol{\sigma}}{\hat{\boldsymbol{v}}_k} - \boldsymbol{f}(\acute{\alpha}_k - \delta_k) \right\|_2, \\ \acute{\alpha}_k - \delta_k, & \text{otherwise.} \end{cases}$$

SPLIT applies a threshold to the initial frequency response parameter, α_1, where if $\acute{\alpha}_1 \notin [-6, 6]$, then the estimation is considered to be not of a scattering center and the estimated value is discarded [11]. SPLIT applies a threshold to δ_k, where if $\delta_k < 0.001$, then α_{k+1} has converged on the prescribed amount of precision and is the finalized estimate of frequency response, α_K [11]. SPLIT then applies a threshold to the finalized estimate of α_K, where if $\alpha_K \notin [-4, 4]$, then the estimation is considered to be not of a scattering center and the estimate value is discarded [11].

2.3.3 Polarization Response Attributes.

The polarization response attributes are extracted by SPLIT from the relationship between the amplitude of the three polarization channels. The characteristics of the physical elements of the target affect the polarization of the field that is re-radiated back to the radar [18]. Specifically, the presence and type of a corner reflector affects the polarization of the re-radiated field. When a linearly polarized electric field is incident on a flat perfect electric conductor (PEC), the reflected field maintains the polarization characteristics of the incident field [18]. When a linearly polarized electric field is incident upon a dihedral corner reflector, the component perpendicular to the reflector becomes inverted [11]. Additionally, for a linearly polarized electric field incident on a trihedral corner reflector, the reflected field maintains the polarization characteristics of the incident field. The nature of the polarization response is documented in Figure 2.5 [11, 18].

Given fully polarimetric SAR data, the polarization response may be extracted. Fully polarimetric SAR data is only attainable with a system pre-configured to transmit, receive, and process radar waveforms with both vertical and horizontal polarization simultaneously. Radar returns consisting of only one polarization cannot be processed to extract a polarization response.

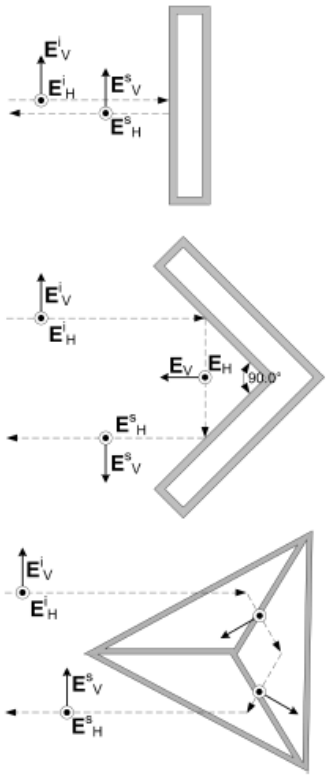

Figure 2.5: Reflection Behavior for Linearly Polarized Electric Fields (used with permission: Dane Fuller) [11, 18].

The relationship between the incident field and the scattered or received field contains the polarization response information. The relationship is defined as [19]

$$\boldsymbol{E}^s = \begin{bmatrix} E_H^s \\ E_V^s \end{bmatrix} = \frac{1}{\sqrt{4\pi r}} \boldsymbol{S}\boldsymbol{E}^i e^{-jkr} = \frac{1}{\sqrt{4\pi r}} \begin{bmatrix} S_{HH} & S_{HV} \\ S_{VH} & S_{VV} \end{bmatrix} \begin{bmatrix} E_H^i \\ E_V^i \end{bmatrix} e^{-jkr}, \tag{2.8}$$

where $\boldsymbol{E}^i$ is the incident electric field, $\boldsymbol{E}^s$ is the scattered or received electric field, r is the distance between the receive antenna and the scatterer, and $\boldsymbol{S}$ is the Sinclair Matrix. The Sinclair Matrix is composed of horizontal and vertical, transmit and receive components. It is important to note the S_{HV} is equivalent to S_{VH} in the case of a monostatic radar [7, 15, 20]. The equivalence of S_{HV} and S_{VH} does not hold true in the case of bistatic radar.

The attributes are decomposed using Krogager decomposition referencing circular polarization. The Sinclair Matrix is translated to circular polarization, where R is right-hand polarized and L is left-hand polarized, by [21]

$$S_{RR} = jS_{HV} + \frac{1}{2}(S_{HH} - S_{VV}), \tag{2.9}$$

$$S_{LL} = jS_{HV} - \frac{1}{2}(S_{HH} - S_{VV}), \tag{2.10}$$

$$S_{RL} = \frac{1}{2}(S_{HH} + S_{VV}). \tag{2.11}$$

From the circular polarization, a measure of the odd, even, and helical scattering mechanisms are extracted by the Krogager decomposition given by [21]

$$k_e = \min(|S_{LL}|, |S_{RR}|), \tag{2.12}$$

$$k_o = |S_{RL}|, \tag{2.13}$$

$$k_h = \text{abs}(|S_{RR}| - |S_{LL}|), \tag{2.14}$$

where k_e is a coefficient of the even bounce mechanism, k_o is a coefficient of the odd bounce mechanism, and k_h is a coefficient of the helical bounce mechanism. The finalized attributes are the normalized coefficient of the bounce mechanisms defined by

$$\kappa_o = \frac{k_o}{\sqrt{|k_o|^2 + |k_e|^2 + |k_h|^2}}, \tag{2.15}$$

$$\kappa_e = \frac{k_e}{\sqrt{|k_o|^2 + |k_e|^2 + |k_h|^2}}. \tag{2.16}$$

SPLIT does not extract the helical bounce mechanism because the helical mechanism can be defined as the relationship between k_e and k_o as [11]

$$k_h = 1 - k_o - k_e. \tag{2.17}$$

We use the k_e and k_o attributes extracted for every pixel in Chapters III and IV. Once we have the extracted amplitude, frequency, and polarization attributes from SPLIT, we form them into features and evaluate the classification performance using machine learning classifiers.

2.4 Classifiers

We used both a linear and a non-linear machine learning classifier to evaluate the classification performance of feature vectors. The linear machine learning classifier we use is linear discriminant analysis (LDA). The non-linear machine learning classifier we use is relevance vector machine (RVM). Each classifier is described in detail in Subsection 2.4.1 and 2.4.2.

2.4.1 Linear Discriminant Analysis.

Linear discriminant analysis (LDA) is a supervised machine learning method for dimensionality reduction, classification, and learning [22]. The process projects a P-dimensional feature vector into a one-dimensional space. The process statistically minimizes the variance of class data in the one-dimensional space, while maximizing separation between classes. LDA is only applicable when P is greater than or equal to 2.

LDA develops a projection, $\mathbf{w}$, such that [22]

$$z = \mathbf{w}^T \mathbf{x}, \tag{2.18}$$

where $\mathbf{x}$ is the feature vector, and z is a point in one-dimensional space. The projection matrix, $\mathbf{w}$, is defined such that the classes are projected to maximize the separation between classes and minimize the scatter within a class [22]. Given a two class comparison with P total features and N instances of each class, there exists the mean of $\mathbf{x_{1,n}} \in \Re^P$, $\mathbf{m_1}$ of class 1 and a mean of $\mathbf{x_{2,n}} \in \Re^P$, $\mathbf{m_2}$ of class 2. There also exists a projection of $\mathbf{m_1}$ and $\mathbf{m_2}$, m_1 and m_2, such that [22]

$$m_1 = \mathbf{w}^T \mathbf{m_1}, \tag{2.19}$$

$$m_2 = \mathbf{w}^T \mathbf{m_2}. \tag{2.20}$$

The scatter within a class, s_1^2 for class 1 and s_2^2 for class 2, is characterized as

$$s_1^2 = \sum_{n=1}^{N} (\mathbf{w}^T \mathbf{x_{1,n}} - m_1)^2, \tag{2.21}$$

$$s_2^2 = \sum_{n=1}^{N} (\mathbf{w}^T \mathbf{x_{2,n}} - m_2)^2. \tag{2.22}$$

The objective is to maximize $|m_1 - m_2|$ and minimize $(s_1^2 + s_2^2)$ [22]. The $\mathbf{w}$ maximizing

$$J(\mathbf{w}) = \frac{(m_1 - m_2)^2}{s_1^2 + s_2^2}, \tag{2.23}$$

is the Fisher's linear discriminant [22]. From the numerator, we derive the between-class scatter matrix, $\mathbf{S}_B$, through [22]

$$\begin{aligned} (m_1 - m_2)^2 &= (\mathbf{w}^T \mathbf{m_1} - \mathbf{w}^T \mathbf{m_2})^2, \\ &= \mathbf{w}^T (\mathbf{m_1} - \mathbf{m_2})(\mathbf{m_1} - \mathbf{m_2})^T \mathbf{w}, \\ &= \mathbf{w}^T \mathbf{S}_B \mathbf{w}. \end{aligned} \tag{2.24}$$

The within-class scatter matrix, $\mathbf{S}_c$ is extracted by rewriting the variance of a class after projection as

$$\begin{aligned} s_c^2 &= \sum_n \mathbf{w}^T (x_{c,n} - \mathbf{m_c})(x_{c,n} - \mathbf{m_c})^T \mathbf{w}, \\ &= \mathbf{w}^T \mathbf{S}_c \mathbf{w}, \end{aligned} \tag{2.25}$$

where subscript $c \in [1, 2]$ is the class designator and $\mathbf{S}_c = \sum_n (x_{c,n} - \mathbf{m_c})(x_{c,n} - \mathbf{m_c})^T$. Substituting $s_1^2 + s_2^2$ in the denominator of Equation (2.23) with

$$\begin{aligned} s_1^2 + s_2^2 &= \mathbf{w}^T \mathbf{S}_1 \mathbf{w} + \mathbf{w}^T \mathbf{S}_2 \mathbf{w}, \\ &= \mathbf{w}^T \mathbf{S}_W \mathbf{w}, \end{aligned} \tag{2.26}$$

where $\mathbf{S}_W = \mathbf{S}_1 + \mathbf{S}_2$, and the numerator with Equation (2.24), Equation (2.23) is rewritten as

$$J(\mathbf{w}) = \frac{\mathbf{w}^T \mathbf{S}_B \mathbf{w}}{\mathbf{w}^T \mathbf{S}_W \mathbf{w}}. \tag{2.27}$$

To evaluate where $J(\mathbf{w})$ is maximized, the gradient with respect to $\mathbf{w}$ is taken and set equal to zero. The result is [22]

$$\frac{\mathbf{w}^T(\mathbf{m}_1 - \mathbf{m}_2)}{\mathbf{w}^T\mathbf{S}_W\mathbf{w}}\left(2(\mathbf{m}_1 - \mathbf{m}_2) - \frac{\mathbf{w}^T(\mathbf{m}_1 - \mathbf{m}_2)}{\mathbf{w}^T\mathbf{S}_W\mathbf{w}}\mathbf{S}_W\mathbf{w}\right) = 0. \tag{2.28}$$

Solving Equation (2.28) for $\mathbf{w}$,

$$\mathbf{w} = \mathbf{S}_W^{-1}(\mathbf{m}_1 - \mathbf{m}_2). \tag{2.29}$$

LDA is optimal when the classes are normally distributed [22]. In such a case, the distribution of class c is $\mathcal{N}(\mathbf{m}_c, \mathbf{S}_W)$, where $\mathbf{S}_W$ is the same as in Equation (2.29). Additionally, if $\mathbf{m}_1 \approx \mathbf{m}_2$, then $\mathbf{w}$ approaches zero as $\mathbf{m}_1$ goes to $\mathbf{m}_2$. In such a case, the classes are inseparable with the features used.

We use the LDA classifier later in Chapter III. We do not expect the LDA to perform well with the high dimensionality of the target states. If $\mathbf{m_1} \approx \mathbf{m_2}$, then LDA is unable to separate the classes. Instead, we use the non-linear classifier, RVM, to classify in the high dimensional space.

2.4.2 Relevance Vector Machine.

Relevance vector machine (RVM) is a supervised machine learning process using a Bayesian framework and kernel functions to obtain sparse solutions to non-linear classification tasks [23]. RVM uses a Bayesian framework applied to the structure of another sparse linearly-parameterized model, the support vector machine (SVM). Similar to LDA, SVM attempts to maximize the spread between classes and minimize the error or variance within a class [24, 25].

SVM classification decisions are based on [24–26]

$$y_i = \text{sgn}(\mathbf{v}^T\mathbf{x}_i + b), \tag{2.30}$$

where $\mathbf{v}$ is a vector of weights defining the hyperplane with a crossing at b, $\mathbf{x}$ is a feature vector, and y_i is the class identifier $y_i \in \{-1, 1\}$. The hyperplane separates the $y_i = 1$

and $y_i = -1$ classes given linearly separable data. The SVM optimizes the hyperplane $\mathbf{v}$, defined by [25],

$$\mathbf{v} = \sum_{i=1}^{l} y_i \alpha_i \mathbf{x}_i, \tag{2.31}$$

where l is the total number of training vectors. The α_i are Lagrange multipliers and the primal Lagrangian is [25, 27],

$$L(\boldsymbol{\alpha}) = \sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{l} y_i y_j \alpha_i \alpha_j \langle \mathbf{x}_i \cdot \mathbf{x}_j \rangle, \tag{2.32}$$

where $\langle\rangle$ is the kernel operator we will discuss later in this section. The kernel in Equation (2.32) is the dot product of x_i and x_j. The vector $\boldsymbol{\alpha}^*$ that maximizes the primal Lagrangian in Equation (2.32) while also holding to [25]

$$\sum_{i=1}^{l} y_i \alpha_i = 0, \tag{2.33}$$

and

$$\alpha_i \geq 0, i = 1, ..., l, \tag{2.34}$$

optimizes the hyperplane $\mathbf{v}$ in Equation 2.31. The optimized hyperplane, $\mathbf{v}^*$ is defined as

$$\mathbf{v}^* = \sum_{i=1}^{l} y_i \alpha_i^* \mathbf{x}_i. \tag{2.35}$$

The value of b^* is defined where $\mathbf{v}^*$ optimally separates the two classes by [25]

$$b^* = -\frac{\max_{y_i=-1} \langle \mathbf{v}^* \cdot \mathbf{x_i} \rangle + \min_{y_i=1} \langle \mathbf{v}^* \cdot \mathbf{x}_i \rangle}{2}. \tag{2.36}$$

The Karush-Kuhn-Tucker complementary conditions apply such that [25, 27]

$$\alpha_i^* [y_i (\langle \mathbf{v}^* \cdot \mathbf{x}_i \rangle + b^*) - 1] = 0, \quad i = 1, ..., l, \tag{2.37}$$

implying that only inputs $\mathbf{x}_i$ closest to the hyperplane have a corresponding non-zero α_i^* [25]. The $\mathbf{x}_i$ with non-zero α_i^* are called support vectors [23–25]. Figure 2.6 illustrates the concept in a two-dimensional linearly separable feature space. The circled points are the support vectors.

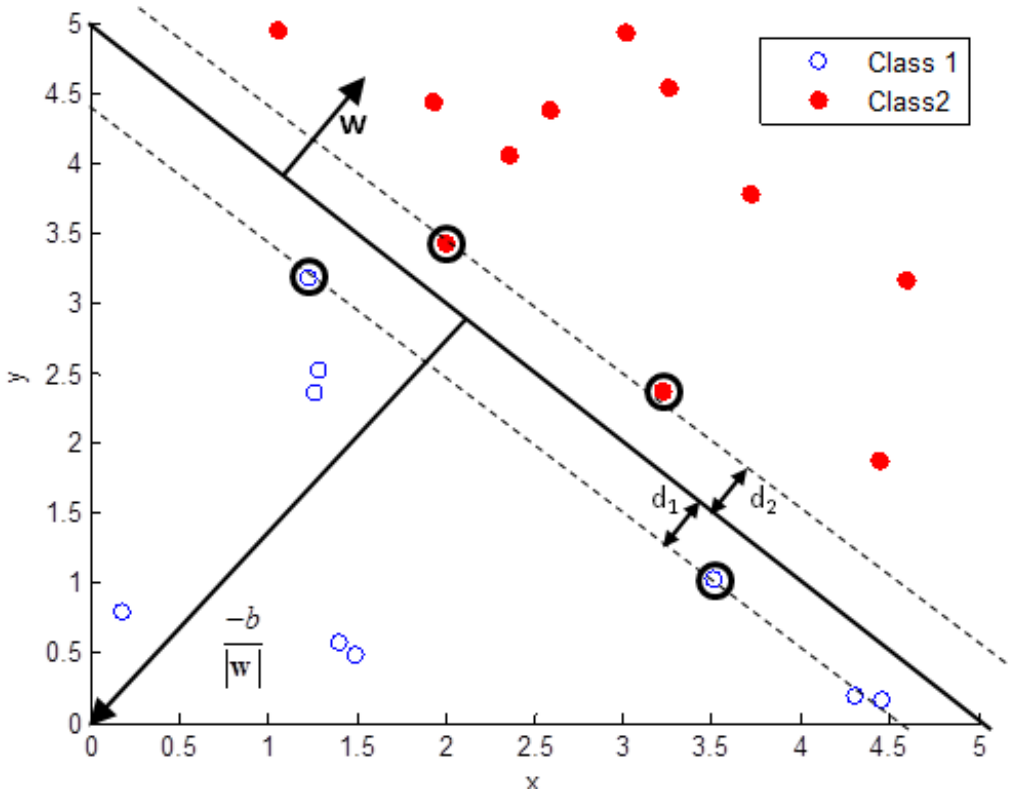

Figure 2.6: Hyperplane Through Two Linearly Separable Classes [24].

Using the support vectors, the optimal hyperplane can be expressed without explicitly defining the hyperplane $\mathbf{v}^*$ as [25]

$$f(\mathbf{x}, \boldsymbol{\alpha}^*, b^*, \mathbf{x}_{\in sv}) = \sum_{i \in sv} y_i \alpha_i^* \langle \mathbf{x}_i \cdot \mathbf{x} \rangle + b^*, \tag{2.38}$$

where sv are the indices of the support vectors. From Equation (2.38), $\langle \mathbf{x}_i \cdot \mathbf{x} \rangle$ is defined as the linear kernel.

For classification of non-linearly separable data, the linear kernel does not provide separation between the classes in a linear feature space. A non-linear kernel is used to implicitly map $\mathbf{x}_i$ into a non-linear feature space. The non-linearly separable classes shown in Figure 2.7 are defined by two parameters, one on each axis in the left image. Notice that a hyperplane cannot be defined to separate the classes. However, the classes are separable in

a third-dimension, as shown on the right side of Figure 2.7. The third dimension is similar to a non-linear feature space created by a non-linear kernel. SVM uses kernel mapping to evaluate in the third dimension. In this thesis we use non-linear kernel called the radial

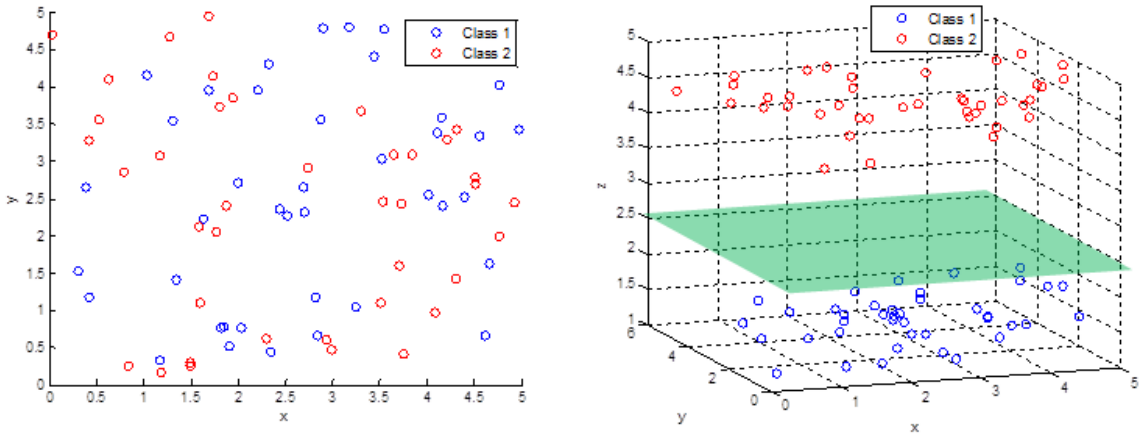

Figure 2.7: Data Re-Mapping Using the Radial Basis Function.

basis function, defined as [24]

$$\mathrm{K}(\mathbf{x}_i, \mathbf{x}_j) = e^{-\left(\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2}\right)}, \tag{2.39}$$

where the subscripts on $\mathbf{x}$ are individual instances of the training feature vectors. The form of the primal Lagrangian to optimize the hyperplane using the non-linear radial basis kernel is [24]

$$L(\boldsymbol{\alpha}) = \sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{l} y_i y_j \alpha_i \alpha_j K(\mathbf{x}_i, \mathbf{x}_j). \tag{2.40}$$

Solving for $\boldsymbol{\alpha}^*$ as in [25], b^* is defined by [24]

$$b^* = \frac{1}{N_s} \sum_{i \in sv} (y_i - \sum_{j \in sv} \alpha_j^* y_j K(\mathbf{x}_i, \mathbf{x}_j)). \tag{2.41}$$

The hyperplane is then defined similar to Equation (2.38) as [24, 25, 27]

$$f(x, \boldsymbol{\alpha}^*, b^*) = \sum_{i \in sv} y_i \alpha_i K(\mathbf{x}_i, \mathbf{x}) + b^*. \tag{2.42}$$

Using the same process as with the radial basis function, other non-linear kernels may be implemented.

SVM develops a hyperplane for use as a classification boundary, given sufficient training vectors. SVM distills the training vectors down to support vectors used to define the hyperplane. The hyperplane is defined using the kernel function in conjunction with support vectors. There exists a linear correlation between the number of training vectors and support vectors [23]. As the number of support vectors grows, so does the number of basis functions of the kernel function. For large training sets, the increase in support vectors becomes computationally prohibitive. Additionally, the increase in the support vectors reduces the smoothness of the boundary between classes leading to over-classification. In response to the faults of SVM, RVM was developed by Tipping [23].

RVM is a Bayesian approach to SVM [23, 28]. The Bayesian approach further increases the sparseness already present in the SVM support vectors by inclining the α_i value to zero [23, 28]. The new set of vectors, called relevance vectors, are sparsely determined with the posterior distribution of the training vectors and a limiting prior distribution on the weight α_i [23].

RVM manipulates the optimal hyperplane in Equation (2.42), removing y_i and b^* to be [23, 28]

$$f(\mathbf{x}, \boldsymbol{\alpha}^*) = \sum_{i \in rv} \alpha_i^* K(\mathbf{x}_i, \mathbf{x}), \tag{2.43}$$

where rv are the indices of relevance vectors. Assuming a Gaussian distribution of y_i, the likelihood function of the complete data set $P(\mathbf{y}|\mathbf{x}, \boldsymbol{\alpha}, \sigma^2)$ is defined as [28],

$$P(\mathbf{y}|\mathbf{x}, \boldsymbol{\alpha}, \sigma^2) = \prod_{i=1}^{N} (2\pi\sigma^2)^{-\frac{1}{2}} e^{(\frac{-1}{2\sigma^2}(y_i - f_i)^2)}, \tag{2.44}$$

where N is the number of training vectors, n and i are the indexes of the training vectors, and

$$f_i = \sum_{n=1}^{N} \alpha_n K(\mathbf{x}_n, \mathbf{x_i}). \tag{2.45}$$

Constraining the magnitude of α, RVM uses a bias in the form of zero-mean Gaussian prior [28],

$$p(\alpha|\vartheta) = \prod_{i=1}^{N} N(0, \vartheta_i^{-1}). \tag{2.46}$$

A Gaussian prior distribution is enforced over the α_i values with a mean of zero [23]. The variance $\boldsymbol{\vartheta}$ is defined through the maximization of the marginal likelihood, lending to more α_i weights evaluating to zero, and resulting in an increasingly sparse relevance vector set [23]. From the prior in Equation (2.46) of $\boldsymbol{\alpha}$ and the likelihood function (2.44), the posterior probability is represented as [23, 28]

$$p(\boldsymbol{\alpha}, \boldsymbol{\vartheta}, \sigma^2|\mathbf{y}) = p(\boldsymbol{\alpha}|\mathbf{y}, \boldsymbol{\vartheta}, \sigma^2)p(\boldsymbol{\vartheta}, \sigma^2|\mathbf{y}). \tag{2.47}$$

From the posterior probability in Equation (2.47), the marginal likelihood $P(\mathbf{y}|\boldsymbol{\vartheta}, \sigma^2)$ is derived [23, 28]. The maximization of the marginal likelihood with respect to $\boldsymbol{\vartheta}$ and σ^2 gives the optimal hyperplane [23, 28].

Implementing the additional Bayesian constraints of RVM on the SVM classification method produces a sparse set of relevance vectors [23]. The sparsity of the relevance vector set, enforced by Equation (2.46), limits the number of basis functions, and smooths the hyperplane. We use RVM to evaluate the classification performance of different feature vectors in Chapters III and IV. The work done by Flynn [7] used RVM to compare classification performance of different combinations of bandwidth, elevation, azimuth and aperture.

2.5 Previous Work With Pixel Attributes From SPLIT

Previous work performed by Flynn used SPLIT to extract feature vectors from pixels of SAR images [7]. Flynn used multiple feature vectors extracted from each image to classify the vehicle in the image [7]. The work in this thesis is a follow-on effort to Flynn's research.

2.5.1 Flynn's Work on Pixel Attributes From SPLIT.

The work performed by Flynn analyzed the impact of bandwidth, elevation, azimuth and aperture on classification performance [7]. The features used for classification were formed directly from peak pixel attributes, $[\alpha, k_e, k_o, x, y]$, extracted using SPLIT. The features from a single pixel made up the feature vector used to classify the image. RVM and the AFRL civilian vehicles (CV) Domes data set was used to evaluate the performance of different combinations of bandwidth, elevation, azimuth, and aperture. The next sections discuss CV Domes and Flynn's work [7] in more detail.

2.5.2 CV Domes.

Training and testing data was drawn from the CV Domes data set. The CV Domes data set is a collection of X-band scattering data for a set of ten vehicles [1]. Fully-polarized far-field monostatic scatter data is simulated over 360 degrees of azimuth and from 30 degrees up to 60 degrees of elevation. In azimuth, data was simulated every 0.0625 degrees, resulting in a total of $5,760$ azimuth samples for each elevation. In elevation, the data was simulated every 0.0625 degrees, resulting in a total of 480 elevation samples per azimuth angle.

Phase history is generated for each of the ten vehicles in Figure 1.1 [1]. For each elevation and azimuth pair, a 1-dimensional profile is simulated with 512 frequency samples with a center frequency of 9.6 GHz and a bandwidth of 5.35 GHz [1]. The 1-dimensional profiles are documented in the frequency domain as phase history. The data is fully polarimetric with HH, HV, and VV linear polarization channels [1].

The CV domes data using full bandwidth with an aperture of 20 degrees has a cross-range resolution of 0.0448 meters and a range resolution of 0.0280 meters. Range resolution, ρ_x, is a function of the bandwidth, B, and the speed of light, c, where [8]

$$\rho_x = \frac{c}{2B}. \tag{2.48}$$

Cross-range resolution, ρ_y, is a function of the wavelength, λ, and the aperture, $\Delta\phi$, in radians where [8]

$$\rho_y = \frac{\lambda}{2\Delta\phi}. \tag{2.49}$$

2.5.3 Flynn's Results.

Flynn's research using a single observation state and feature vectors tied back to a single pixel classified SAR images [7]. Flynn concluded the research with a recommendation on the bandwidth, elevation, azimuth, and aperture collection parameter for SAR images. He used the CV Domes data set to simulate classification performance with different bandwidth, elevation, azimuth, and aperture parameters. From the classification performance results, Flynn identified a parameter set with the highest performance [7]. Flynn recommends an azimuth angle of 90 degrees to 135 degrees, an aperture size of 60 degrees, an elevation angle of 30 degrees, and a bandwidth from 640 MHz to 3 GHz based on the performance of the parameters [7]. Similarly, we compare feature vectors in Chapters III and IV.

2.6 Research Goals and Assumptions

We want to implement the pixel method from [7] using the full extent of observation angles and an aperture of 20 degrees. Considering the concept of operation of a SAR platform [29], we conclude the platform has limited control of the observation angles of a target. Because there is limited control of the elevation and azimuth of a target, classification performance must be evaluated using the full extent of observation angles. Additionally, given a bandwidth of 5.35GHz and a center frequency f_c = 9.6GHz, a 60 degree aperture is considered a wide-angle synthetic aperture [30]. A wide-angle synthetic aperture is any synthesized aperture having an angular extent, $\Delta\phi$, greater than required to have equivalent resolution in range and cross-range given by [30]. A wide-angle synthetic aperture is defined as [30]

$$\Delta\phi > 2\sin^{-1}\left(\frac{BW}{2f_c}\right), \tag{2.50}$$

where BW is the bandwidth and f_c is the center frequency. Also, unlike ideal point scatters, canonical scatterers have an angular persistence of less than 20 degrees [31]. For these reasons, we use a 20 degree aperture. Note that, given a distinct SAR collection scenario, the aperture may be different, as the aperture is dependent on the equipment and the concept of operation.

From the CV domes data set, there are a total of 27,705,600 possible target states given the constraints of a 20 degree aperture, and a constant elevation angle across the aperture. Figure 2.8 illustrates all possible target states. Separating the CV Domes into sedans and SUVs, there are 16,623,360 possible sedan states and 11,082,240 possible SUV states. To reduce the data size and respect computational limitations, we sparsely sample from the full extent of elevation and azimuth. The rate at which we sample the target states is defined in Sections 3.1 and 3.3.

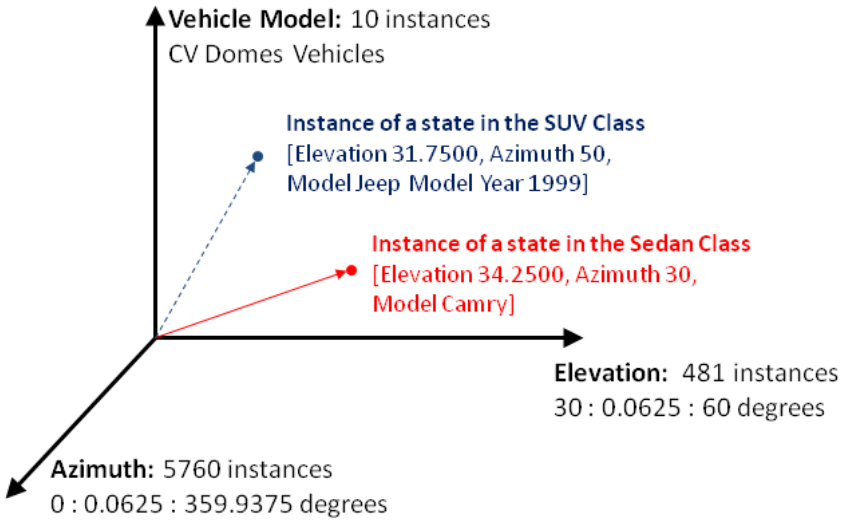

Figure 2.8: All Possible Target States From the CV Domes Data Set.

We want to capture the entire image in a single feature vector to investigate the attributes of the entire image for impact on classification performance. Evaluating the performance of subsets of features in the feature vector enables us to relate classification performance to individual features. If the features, and the attributes they are composed of, are tied back to physical elements of the vehicles, then we can tie physical elements of the vehicles to impact in classification performance.

In Chapter III, we construct a process to form feature vectors from an entire SAR image. First, we extend the research by Flynn [7] to use more target states and investigate the corresponding classification performance of subsets of the overall feature vector. With a baseline of the performance from Flynn's feature extraction method [7], we extend the method to form features across the entire image space. Utilizing features from the entire image space, we investigate the corresponding classification performance of subsets of all the features we extract from the image.

III. Application of Feature Extraction Methods

The goal of this chapter is to apply various methods for feature extraction and to evaluate their corresponding classification performance. Classification performance is evaluated using the linear and non-linear classifiers, LDA and RVM respectively. In Sections 3.1 and 3.2, the classification performance of feature vectors formed using the pixel method is evaluated. In Sections 3.3-3.5, the classification performance of feature vectors formed using the cell method is evaluated. Section 3.6 reports the notable conclusions from the pixel method and the cell method. Section 3.7 introduces the concept of feature saliency.

3.1 Pixel Method for Extracting Features

The pixel method for extracting features is similar to the method used by Flynn [7]. There are two major differences between the implementation of the pixel method employed in this work and previous work [7]. First, we train and test using target states spanning 180 degrees in azimuth and up to nine degrees in elevation. The second difference is the implementation of a segmentation process to reduce the number of feature vectors extracted from each target state's image.

3.1.1 Pixel Method.

We use the process shown in Figure 3.1 to implement the pixel method. Attributes are extracted from the data using SPLIT. Features are formed from the attributes of individual pixels. The extracted features then are used to populate a feature vector to be analyzed. The classification performance of various feature vectors is then simulated and evaluated. The image segmentation and formation of features shown in the third block of Figure 3.1, is unique to the pixel method. We are unable to process all of the data from the CV Domes data set and pull all possible feature vectors from each image due to computational limits.

We limit the number of feature vectors by sparsely sampling the target states from the CV Domes data set and using image segmentation.

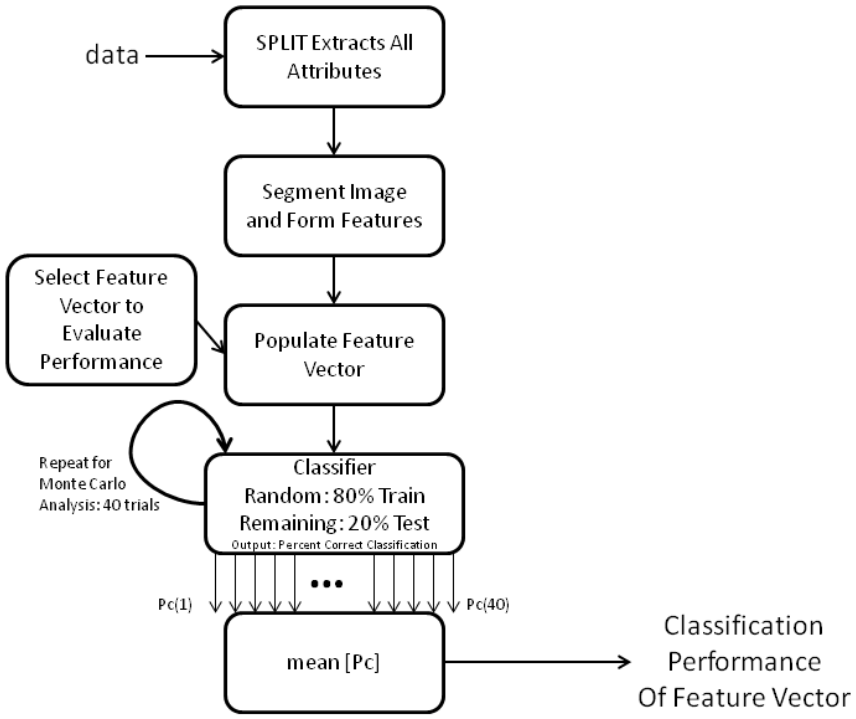

Figure 3.1: Block Diagram of Classification Performance Analysis of Pixel Method Feature Vectors.

Using the extracted feature vectors from all of the pixels with a valid alpha value, as was performed by Flynn in [7], is computationally expensive when implementing multiple observation states. To reduce the computational costs, we reduce the number of feature vectors we extract per image using segmentation. The pixels in the SAR images are segmented using threshold values of all the attributes of pixels, [x, y, A, α, k_o, k_e] based on the ideal mapping shown in Figure 3.2. Unlike previous work [11], we implement a threshold on distances from peak amplitude pixels within the image instead of ideal pixel attributes to segment pixels.

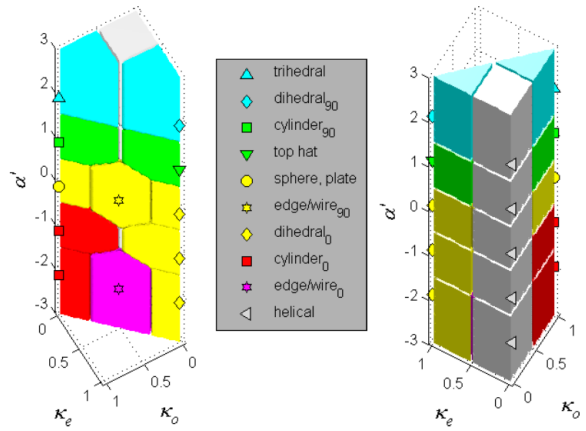

Figure 3.2: Three Dimensional Attribute Space [11].

We utilize the process shown in Figure 3.3 to segment the pixels in an image. The pixels are sorted from highest to lowest by amplitude. The first pixel is defined to be an associate pixel. The subsequent pixels are segmented with an associate pixel or are assigned to be a new associate pixel based on a threshold on distance between pixels in attribute space.

Figure 3.2 shows there are three general cases for the threshold on distance in attribute space under which two pixels can still be classified as the same type of canonical shape. The first case is the ideal mapping of a dihedral$_0$ in attribute space. The second case is the mapping of a cylinder$_0$ in attribute space. The third case is the mapping of the remaining canonical shapes. To model the different cases, we use three different sets of thresholds depending on the location of the associate pixel in attribute space.

- **Case 1:** If the associate pixel is $\alpha_{as} < 0.5$, a $k_{e\ as} > 0.5$, and a $k_{o\ as} < 0.5$.
- **Case 2:** If the associate pixel is $\alpha_{as} < -0.5$, a $k_{e\ as} < 0.5$, and a $k_{o\ as} < 0.5$.
- **Case 3:** If the associate pixel is not case 1 or case 2.

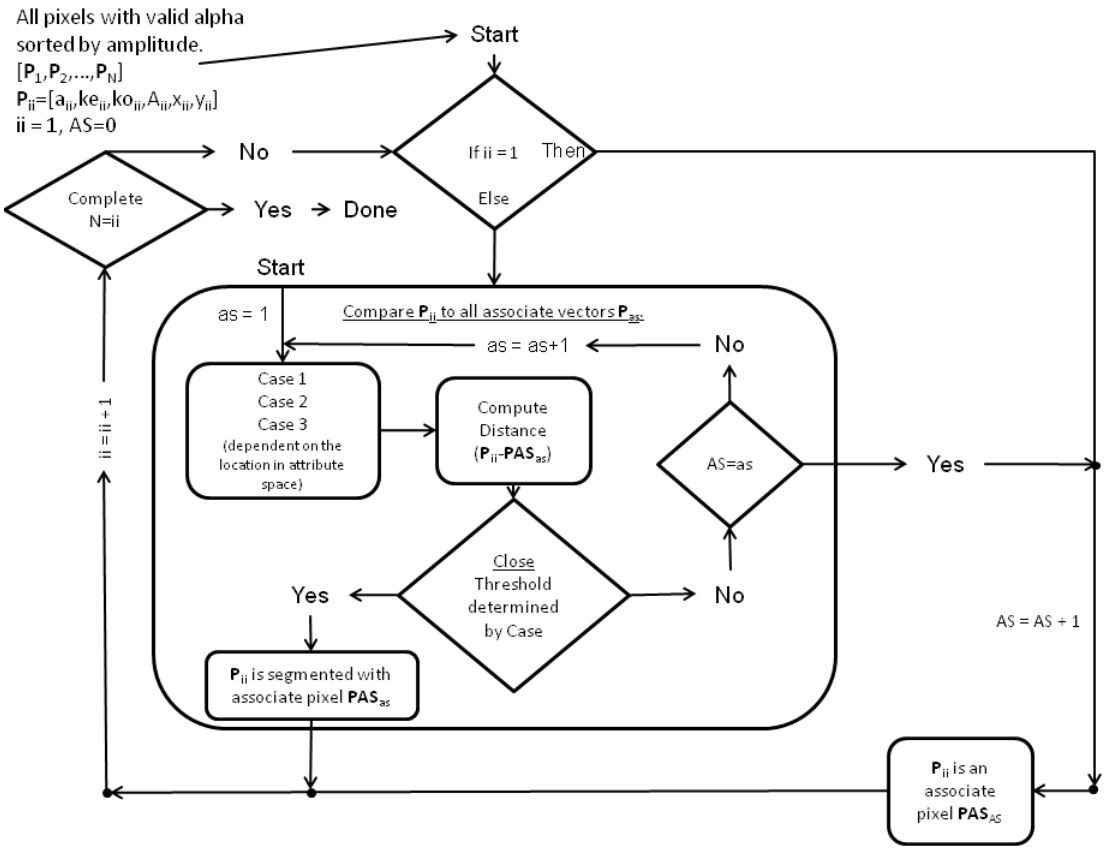

Figure 3.3: Block Diagram of the Image Segmentation Process.

For associate pixels that exist in **Case 1**, the threshold on the distance between a pixel and the associate pixel is

$$|\alpha_{as} - \alpha_{ii}| \leq 3 \;\&\; |k_{e\,as} - k_{e\,ii}| \leq 0.5 \;\&\; |k_{o\,as} - k_{o\,ii}| \leq 0.5 \;\&\; |x_{as} - x_{ii}| \leq 0.75 \;\&\; |y_{as} - y_{ii}| \leq 0.75, \tag{3.1}$$

where the subscript as is the index of associate pixels and the subscript ii is the index of the pixel being segmented. For associate pixels that exist in **Case 2**, the threshold on the distance between a pixel and the associate pixel is

$$|\alpha_{as} - \alpha_{ii}| \leq 2 \;\&\; |k_{e\,as} - k_{e\,ii}| \leq 0.5 \;\&\; |k_{o\,as} - k_{o\,ii}| \leq 0.5 \;\&\; |x_{as} - x_{ii}| \leq 0.75 \;\&\; |y_{as} - y_{ii}| \leq 0.75. \tag{3.2}$$

For associate pixels that exist in **Case 3**, the threshold on the distance between a pixel and the associate pixel is

$$|\alpha_{as} - \alpha_{ii}| \leq 1 \;\&\; |k_{e\,as} - k_{e\,ii}| \leq 0.5 \;\&\; |k_{o\,as} - k_{o\,ii}| \leq 0.5 \;\&$$
$$|x_{as} - x_{ii}| \leq 0.75 \;\&\; |y_{as} - y_{ii}| \leq 0.75. \tag{3.3}$$

Figure 3.4 shows an example of pixel segmentation based on the outlined process. After

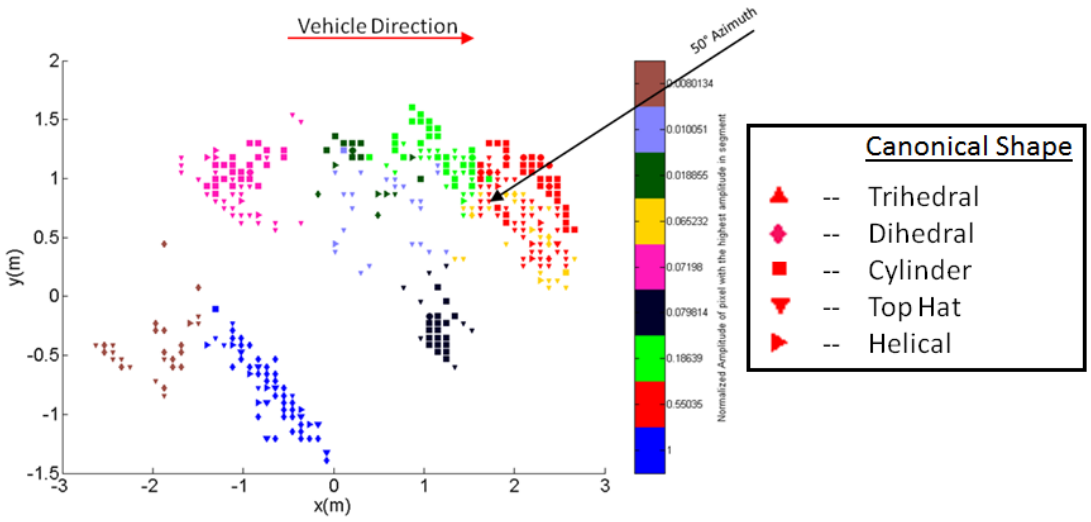

Figure 3.4: Segmented Pixels With the Ideal Mapping of Extracted Attributes to Canonical Shapes of Toyota Camry at 30 Degrees Elevation, and 50 Degrees Azimuth.

the segmentation process feature vectors are formed from the attributes of the associate pixels. The segmentation process shown in Figure 3.3 reduces the number of feature vectors extracted for an image by more than an order of magnitude.

The process of segmentation implicitly captures the pixel amplitude in the location features, x and y. The pixel amplitude is also explicitly captured in the amplitude feature, A. However, we define segments by the pixel with the greatest amplitude, and the location features also contain pixel amplitude information. With the feature vector from

segmentation, we evaluate the classification performance of the feature vectors formed from different sets of features itemized in Table 3.1.

Table 3.1: Pixel Based Feature Vectors.

Reference #	Feature Vector
1	$[\, x,\, y]$
2	$[\, x,\, y,\, A]$
3	$[\, x,\, y,\, k_o]$
4	$[\, x,\, y,\, k_e]$
5	$[\, x,\, y,\, k_o,\, k_e]$
6	$[\, x,\, y,\, A,\, k_o,\, k_e]$
7	$[\, x,\, y,\, \alpha]$
8	$[\, x,\, y,\, A,\, \alpha]$
9	$[\, x,\, y,\, k_o,\, \alpha]$
10	$[\, x,\, y,\, k_e,\, \alpha]$
11	$[\, x,\, y,\, k_o,\, k_e,\, \alpha]$
12	$[\, x,\, y,\, A,\, k_o,\, k_e,\, \alpha]$

If we evaluate with all possible target states from the CV Domes data set, we are unable to process all the data with the computational resources available. Respecting computational limitations, we use the states shown in Figure 3.5. We increase the span of elevation states from 30 to 32 degrees up to 30 to 39 degrees and report the impact of increasing the diversity in elevation to classification performance. Computational limitations prohibit examining a greater diversity in elevation.

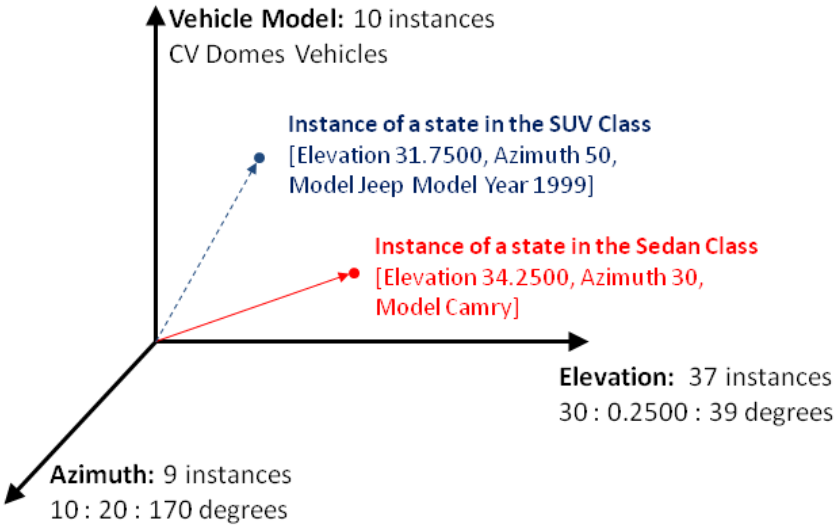

Figure 3.5: Target States Used to Evaluate Classification Performance of Pixel Method.

3.2 Pixel Method Results

We evaluated the feature vectors in Table 3.1 using the process shown in Figure 3.1. The metric used to evaluate classification performance is the mean classification performance for the feature vectors based on 30 trials using randomly chosen training and testing data. We compare the averaged corresponding performance of each feature vector to evaluate the pixel method for classifying targets.

Feature vectors from Table 3.1 containing the amplitude feature consistently have the performance of a 1R classifier [32], where all test feature vectors are classified to be the class with the greatest number of training feature vectors. The classification results for all feature vectors with an amplitude feature have 100 percent correct classification for sedans and zero percent correct classification for SUVs. The classification performance of the remaining feature vectors is reported in Figure 3.6.

The use of polarization response attributes is associated with the highest classification performance. The highest performing feature vectors are feature vectors three and five from

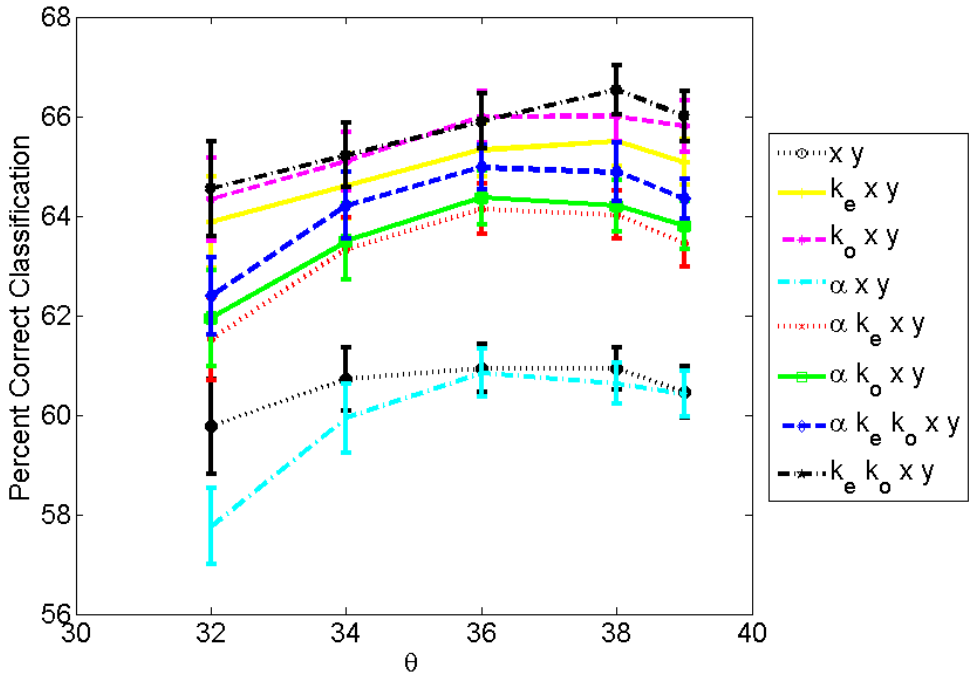

Figure 3.6: Classification Performance Versus Elevation Sampling Diversity. Elevation sampling diversity is from 30 to θ degrees elevation. The error bars are the standard deviation of the classification performance.

Table 3.1, both of which include the odd bounce polarization attribute, k_o. Feature vector four is the next best performing feature vector and includes the even bounce polarization attribute, k_e.

The use of frequency response attributes reduces classification performance. For example, the classification performance of feature vector one is reduced with the addition of the frequency response attribute, α, in feature vector seven. The highest performing feature vector is reduced by more than its standard deviation with the addition of α in feature vector 11.

The classification performance using the pixel method for feature extraction peaks at 66 percent correct classification, although the results from other methods show that higher classification performances are attainable [3–6]. Due to the poor classification performance

of the pixel method under the given test conditions, we want to explore another option for feature extraction.

3.2.1 Motivation for Cell Method.

It is desirable to improve classification performance, and to identify attributes and features which have the greatest impact on the classification performance. Backward selection is a common method for analyzing the impact of features on classification, and will be discussed in Section 4.1. However, the structure of a pixel method feature vector is not appropriate for backward selection. A single feature vector that captures the information within an entire SAR image is required for backward selection. One method of constructing a single feature vector from an image is to use cells as a framework to extract features.

3.3 Cell Method for Extracting Features

An image is composed of a matrix of pixels which may be divided into spatial regions defined as cells and blocks. Cells do not overlap, and four cells compose a spatial region defined as a block. The relationship of a pixel to a cell to a block is captured in Figure 3.7. The use of cells as a framework to extract features is presented in Dalal and Triggs' histograms of oriented gradients (HOG) work [33]. Features are extracted in two ways from the cells. First, the mean and mode of attributes within the cells are used to form features. Second, HOG is used to form features.

Six feature types are formed using the mean and mode of an attribute within a cell. The "α mean" feature type is the mean of the α attribute within a cell. The "α mode" feature type is the mode of the α attributes within a cell. The "k_e mean" feature type is the mean of the k_e attribute within a cell. The "k_e mode" feature type is the mode of the k_e attribute within a cell. The "k_o mean" feature type is the mean of the k_o attribute within a

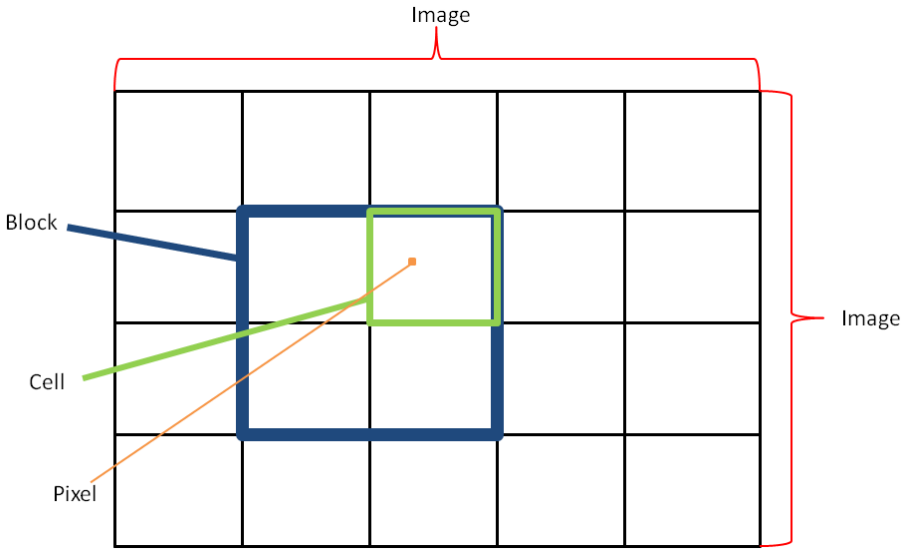

Figure 3.7: The Pixel to Cell to Block Relationship.

cell. The "k_o mode" feature type is the mode of the k_o attribute within a cell. The second method we use to form features is HOG on amplitude and polarization images.

Histogram of oriented gradients (HOG) is a tool originating from computer vision and image processing which defines the orientation of contours within an image space using gradient computations based on cells of an image [33]. HOG calculates the gradient vector, also known as an image gradient, for each pixel, p_c within each cell of an image [33]. The gradient vector of pixel p_c is

$$[\Delta p_x, \Delta p_y] = [p_r - p_l, p_u - p_d], \tag{3.4}$$

where the notation is shown in Figure 3.8. The magnitude of the gradient vector of a pixel, p_{mag}, is $p_{mag} = \sqrt{\Delta p_x^2 + \Delta p_y^2}$. The angle of the gradient vector of a pixel, p_{ang}, is $p_{ang} = \tan^{-1}(\frac{|\Delta p_y|}{|\Delta p_x|})$.

A histogram of gradients is constructed for each cell from the gradient vectors of all of the pixels in each cell. The bins of the histogram are the gradient vector angles. The

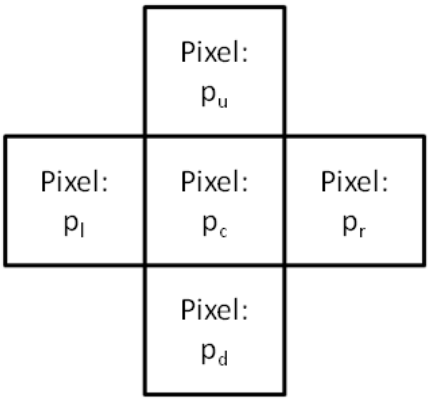

Figure 3.8: Example for Pixel Gradient Vectors.

magnitude in each bin is the sum of the gradient vector magnitudes of each pixel in the bin. From the image in Figure 3.9, HOG forms the histogram in Figure 3.10 for a single cell.

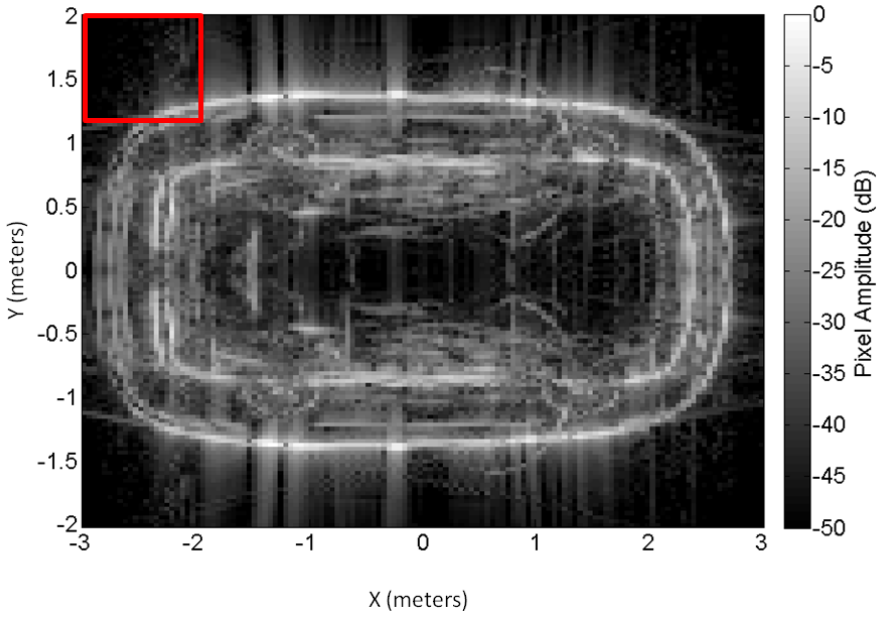

Figure 3.9: Non-Coherently Formed 360 Degree Image of a Toyota Camry, Formed with 20 Degree Apertures.

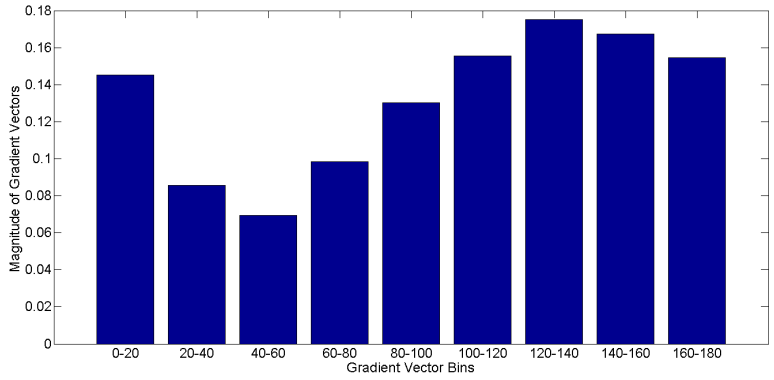

Figure 3.10: Histogram of gradients of a cell of Figure 3.9. Cell spans $x = [-3 : -2]$ and $y = [1.333 : 2]$.

HOG groups cells into larger spatial regions defined as blocks [33]. The blocks overlap, covering the entire image. Within each block, the histograms of the all the cells are normalized, and the normalized copies of the histograms are defined as features. Because of the overlapping blocks, multiple normalized histograms are defined as features for each cell. If HOG is implemented without normalization, then the original histogram bin magnitudes from each cell are also recorded as features. We implement HOG using MATLABs "computer vision" toolbox [34].

Using HOG, we derive three feature types; HOG of amplitude, HOG of k_e, and HOG of k_o. The name of each types comes from the attribute used to develop the image on which HOG operates. Shown in Figure 3.11 is an image developed from the amplitude attributes of pixels extracted using SPLIT. Figure 3.12 shows an image developed from the k_e attributes of pixels extracted using SPLIT. Figure 3.13 shows an image developed from the k_o attributes of pixels extracted using SPLIT.

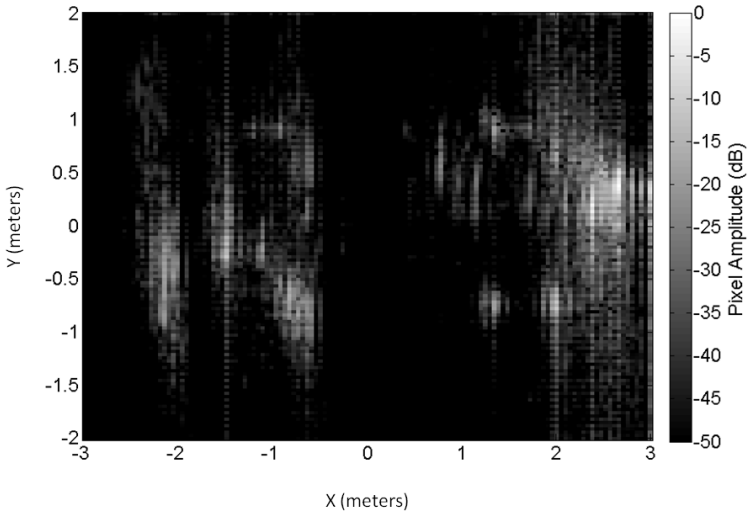

Figure 3.11: Image From Amplitude Attribute of Pixels. Source is a 20 degree aperture of a Toyota Camry at 30 degrees elevation and 10 degrees azimuth.

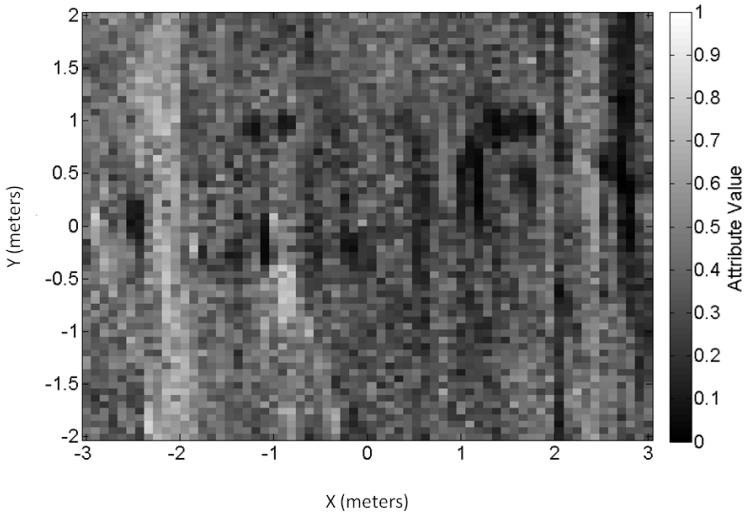

Figure 3.12: Image from k_e Attribute of Pixels. Source is a 20 degree aperture of a Toyota Camry at 30 degrees elevation and 10 degrees azimuth.

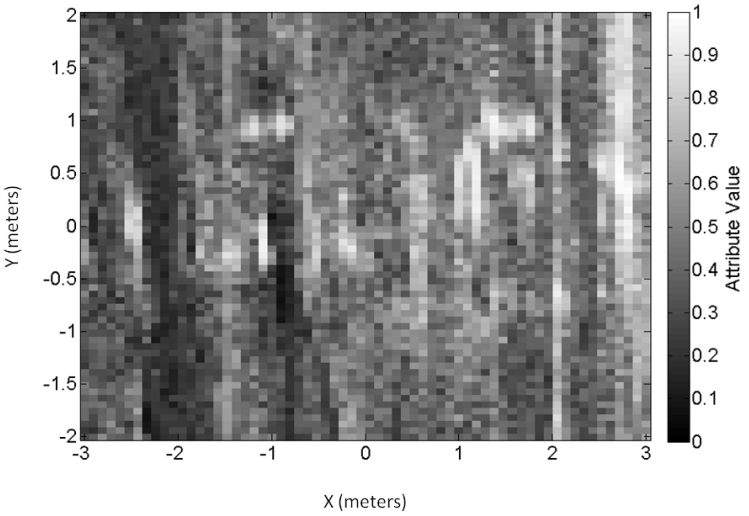

Figure 3.13: Image from k_o Attribute of Pixels. Source is a 20 degree aperture of a Toyota Camry at 30 degrees elevation and 10 degrees azimuth.

3.4 Cell Method Evaluation

We use the process shown in Figure 3.14 to evaluate the classification performance of the cell method. Attributes are extracted from the data using SPLIT. Features are then formed from the attributes of the cells. The resulting features are used to populate feature vectors for evaluation. The classification performance of the resulting feature vectors may then be evaluated.

We evaluate the cell method using a 6x6 grid of cells. A comparison of a 5x5 grid, a 6x6 grid, a 7x7 grid, and an 8x8 grid, shown in Figure 3.15, gives no clear indication of a superior grid size. The 6x6 grid of cells gives a similar symmetric grid of cells and a smaller overall feature vector than an 8x8 grid of cells. The 6x6 grid of cells is laid out and labeled as shown in Figure 3.16. The labels are used for reference in Section 4.3. The SAR image, on which the 6x6 grid is overlaid, is a single target state of a Toyota Camry at 10 degrees azimuth and 30 degrees elevation.

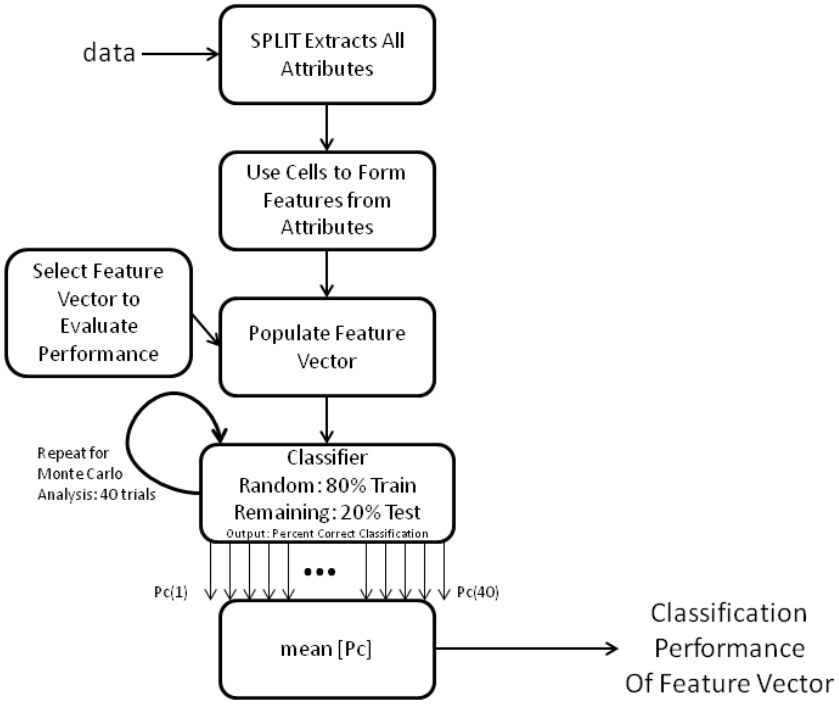

Figure 3.14: Block Diagram of Classification Performance Analysis of Cell Method Feature Vectors.

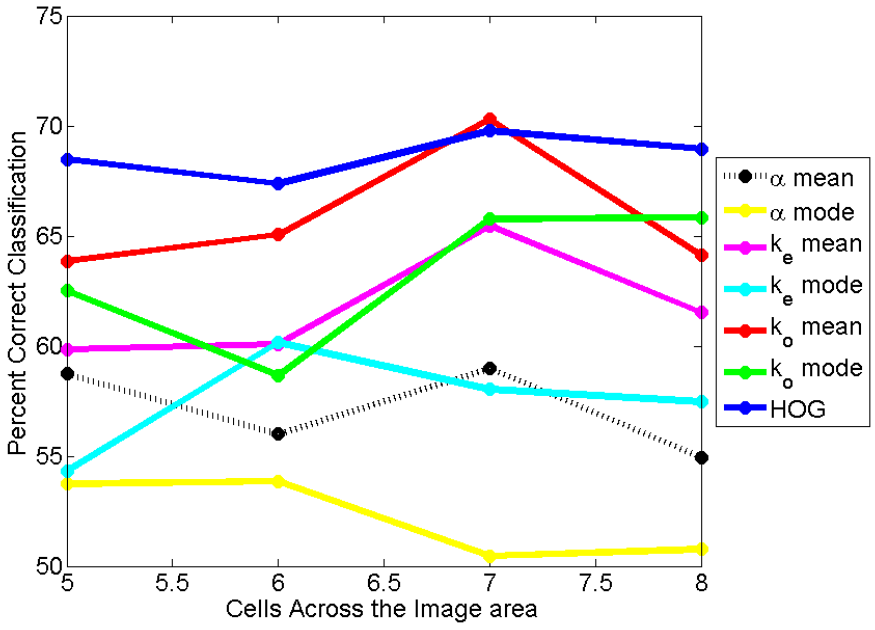

Figure 3.15: Comparison of Cell Size Classification Performance with the RVM Classifier.

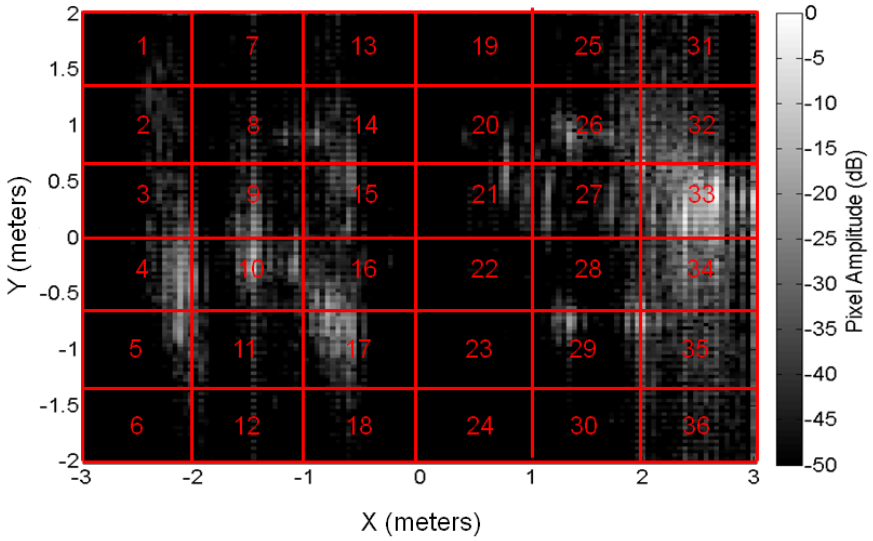

Figure 3.16: 6x6 Grid of Cells Laid Over a SAR Image.

If we evaluate the cell method with all possible target states from the CV Domes data set, then we are unable to process all the data with the computational resources available. To reduce the computational resources required, the CV Domes data set is sparsely sampled every 0.25 degrees in elevation, 20 degrees in azimuth, and using every vehicle. The result is non-overlapping apertures covering 360 degrees on the vehicles and 30 degrees in elevation. A total of 21,780 target states are used. Respecting the computational limit, we use the states shown in Figure 3.17.

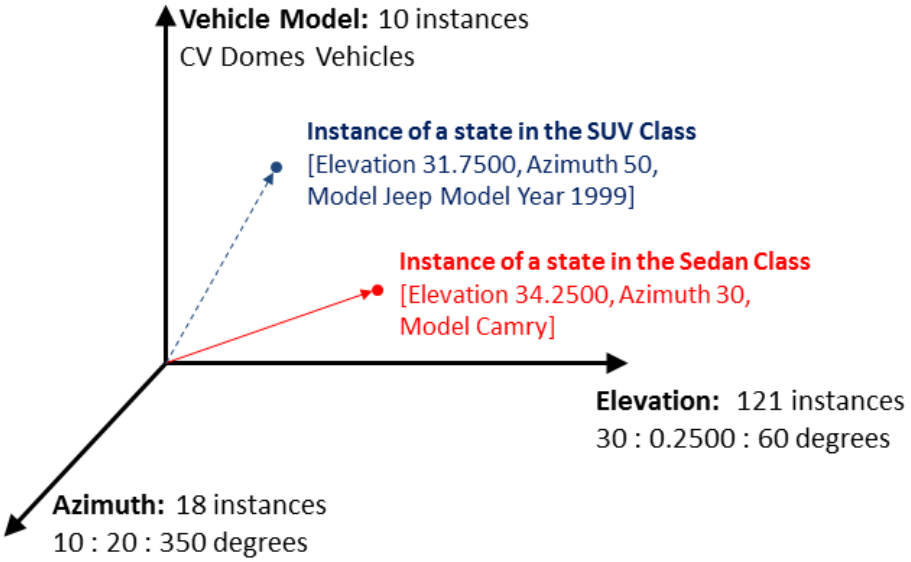

Figure 3.17: Target States Used to Evaluate Classification Performance of Cell Method.

The feature vectors itemized in Table 3.2 are 21 different combinations of feature types we evaluate for classification performance. The vector length column in Figure 3.2 with two entries correspond to the feature vector length with HOG normalization, and without HOG normalization respectively. Comparison of the classification performance of each feature vector shows the comparative impact of the feature types. Section 3.5 reports the classification performance of each feature vector.

Table 3.2: Feature Vectors for Classification.

Feature Vector	α mean	α mode	k_e mean	k_e mode	k_o mean	k_o mode	amp HOG	k_e HOG	k_o HOG	Vector Length
1	√									36
2			√							36
3					√					36
4		√								36
5				√						36
6						√				36
7							√			324, 900
8								√		324, 900
9									√	324, 900
10			√		√					72
11				√		√				72
12	√		√		√					108
13		√		√		√				108
14	√						√			360, 936
15		√					√			360, 936
16			√		√		√			396, 972
17				√		√	√			396, 972
18	√		√		√		√			432, 1008
19		√		√		√	√			432, 1008
20	√	√	√	√	√	√	√			540, 1116
21	√	√	√	√	√	√	√	√	√	1188, 2916

3.5 Cell Method Results

Using the cell method described in Section 3.3, the classification performance of the feature vectors is evaluated and then itemized in Tables 3.3-3.6. The classification performance using the cell method is better than that of the pixel method. Similar to the pixel method results, feature vectors including polarization attributes outperform those which included only amplitude attributes, as well as those including frequency attributes. In some cases, the results also show an enhanced classification performance when a set of feature types are used to construct a feature vector. The classification performance of the LDA classifier proved to be lower than the classification performance of the RVM classifier under the test conditions.

3.5.1 Cell Method Results Using Linear Discriminant Analysis.

The LDA classification performance using the features from the cell method without normalized HOG is reported in Table 3.3 and varies between 56.75 percent correct classification and 75.91 percent correct classification. The use of frequency response features is correlated with the lowest classification performances, which is consistent with the pixel method results. Feature vector one, with only the α mean type of features, shows the lowest classification performance of all the feature vectors. The second worst classification performance is demonstrated by feature vector four, which has only the α mode feature. Out of all the feature types, the α mean and α mode features have the smallest positive impact on classification performance.

The use of the odd bounce polarization response corresponds to the highest classification performance for features not derived using HOG. Feature vector six, with only the k_o mode type of features, shows the highest classification performance out of feature vectors one through nine, which have only one feature type. Feature vector three, with only the k_o mean type of features, has the second highest classification performance.

Table 3.3: Feature Vectors Classification Results Using LDA Without HOG Normalization.

Feature Vector	α mean	α mode	k_e mean	k_e mode	k_o mean	k_o mode	amp HOG	k_e HOG	k_o HOG	Performance % Correct
1	√									56.75
2			√							60.58
3					√					62.74
4		√								57.39
5				√						62.66
6						√				63.93
7							√			75.91
8								√		72.77
9									√	74.21
10			√		√					66.29
11				√		√				67.04
12	√		√		√					68.11
13		√		√		√				69.28
14	√						√			72.10
15		√					√			72.33
16			√		√		√			69.84
17				√		√	√			67.58
18	√		√		√		√			69.89
19		√		√		√	√			66.42
20	√	√	√	√	√	√	√			67.09
21	√	√	√	√	√	√	√	√	√	65.62

Using normalization, Dalal and Triggs' improved the classification performance of HOG features for human detection by four percent [33], and similar performance gains are

observed in the comparison of Table 3.4 and Table 3.3 for the HOG features. The use of HOG normalization on the HOG feature types corresponds to an improved classification performance of seven to ten percent correct classification. The additional dimensions in the feature vector, due to the normalization process, improve the linear classification performance of feature vectors with HOG feature types.

Table 3.4: Feature Vectors Classification Results Using LDA With HOG Normalization.

Feature Vector	α mean	α mode	k_e mean	k_e mode	k_o mean	K_o mode	amp HOG	k_e HOG	k_o HOG	Performance % Correct
7							√			85.94
8								√		79.99
9									√	83.67
14	√						√			76.66
15		√					√			74.09
16			√		√		√			69.32
17				√		√	√			67.31
18	√		√		√		√			70.33
19		√		√		√	√			66.64
20	√	√	√	√	√	√	√			66.71
21	√	√	√	√	√	√	√	√	√	65.73

3.5.2 Cell Method Results Using Relevance Vector Machine.

The RVM classification performance varies between 72.65 percent correct classification and 95.90 percent correct classification using the features from the cell method, without normalized HOG, as reported in Table 3.5. The use of frequency response features correlates to the lowest classification performances, which is consistent with the pixel

method results. Feature vector one, with only the α mean feature type, shows the lowest corresponding classification performance of all the feature vectors. The second worst classification performance corresponds to feature vector three, with only the α mode feature type. The inclusion of α features to the highest performing feature vector, feature vector 17, correlates to a reduction in classification performance of 1.58 percent in feature vector 19. A reduction in the classification performance is reported in every instance where an α feature type is added to a feature vector in Table 3.5. Out of all the feature types, the α mean and α mode features have the least positive impact on classification performance.

The use of the odd bounce polarization response corresponds to the highest classification performance. Of all the feature vectors with a single feature type, feature vector three, with only k_o mean feature type, shows the highest corresponding classification performance. The inclusion of k_o feature types always improves the classification performance of the feature vector.

The use of multiple feature types corresponds to the highest classification performances. Feature vector 17, with k_e mode, k_o mode, and HOG feature types, shows the highest overall corresponding classification performance. Feature vector 16, with k_e mean, k_o mean, and HOG feature types, shows the second highest overall corresponding classification performance.

The improvement of classification performance with HOG normalization is reported in Table 3.6. The use of HOG normalization on the HOG feature types corresponds to a minimal improvement in classification performance. Unlike the improvements reported in Table 3.4 using LDA, the use of normalization with HOG reported in Table 3.6 has at best an improvement of 4 percent correct classification. Some of the feature vectors had a decrease in classification performance with the use of the normalization process of HOG (e.g. feature vectors 8, 9, 16, and 17).

Table 3.5: Feature Vector Classification Results Using RVM Without HOG Normalization.

Feature Vector	α mean	α mode	k_e mean	k_e mode	k_o mean	k_o mode	amp HOG	k_e HOG	k_o HOG	Performance % Correct
1	√									72.65
2			√							89.60
3					√					92.32
4		√								72.30
5				√						83.15
6						√				87.12
7							√			91.22
8								√		83.73
9									√	86.10
10			√		√					92.76
11				√		√				94.29
12	√		√		√					87.11
13		√		√		√				89.21
14	√						√			91.11
15		√					√			90.78
16			√		√		√			95.49
17				√		√	√			95.90
18	√		√		√		√			94.23
19		√		√		√	√			94.32
20	√	√	√	√	√	√	√			94.98
21	√	√	√	√	√	√	√	√	√	94.65

Table 3.6: Feature Vector Classification Results Using RVM With HOG Normalization.

Feature Vector	α mean	α mode	k_e mean	k_e mode	k_o mean	k_o mode	amp HOG	k_e HOG	k_o HOG	Performance % Correct
7							√			91.56
8								√		78.81
9									√	83.27
14	√						√			92.43
15		√					√			91.99
16			√		√		√			94.52
17				√		√	√			93.89
18	√		√		√		√			94.60
19		√		√		√	√			94.37
20	√	√	√	√	√	√	√			95.71
21	√	√	√	√	√	√	√	√	√	96.10

3.6 Conclusions

We report a similar trend in classification performance for both LDA and RVM. The use of frequency response features uniformly reduced classification performance. The use of polarization response features and the use of HOG features correlates to the best performance of the feature vectors evaluated. Notably, the combination of feature types in feature vector 17 corresponds to the best performance of all the feature vectors evaluated.

3.7 Saliency of Features

A feature vector formed from a subset of all features results in the highest classification performance. We ask three question. Is there a subset of features with optimal separation between classes? If there is a optimal subset of features, then how do we identify the optimal subset of features? If we identify an optimal subset, then what can we learn from it?

The subset of features whose corresponding classification performance is optimal, compared to all the permutations of the overall features set, may be taken as the salient set of features. Saliency of feature types is measured by the change in classification performance from the removal of the features from the feature type. Chapter IV evaluates the saliency of the cell method features using the process of backward selection.

IV. Application of Backward Selection on Cell Method Features

The goal of Chapter IV is to evaluate the saliency of features using cell method features from Section 3.4 and the CV Domes data set. Section 4.1 introduces backward selection for feature subset selection. In Section 4.2, backward selection is applied to the cell method features. Section 4.3 reports the results of backward selection on the cell method features.

4.1 Backward Selection

Backward selection is a process for subset selection. Subset selection is a process of finding the smallest number of dimensions of a feature set that contribute the most to the accuracy of the classifier [22]. Backward selection starts with all available features and removes them one by one. Within each iteration, the candidate feature whose removal decreases the classification error the most is left out of the feature subset on the next iteration of backward selection [22]. Backward selection iterates and is complete when the removal of features no longer reduces the error in classification.

Backward selection is computationally expensive, but systematically converges on a salient set of features. For each feature removed, the classification performance must be evaluated for each of the remaining features [22]. To reduce the dimension of the set of features from (N) to $(N - r)$ features, classification performance must be evaluated $N + (N - 1) + (N - 2) + \cdots + (N - r + 2) + (N - r + 1)$ times [22]. Backward selection is more efficient than completing a grid search of all permutations of the feature set, which would require $\frac{N!}{(N-r)!r!}$ evaluations.

4.2 Implementation of Backward Selection

4.2.1 Rules.

To apply backward selection to the cell method feature vectors, the following rules are adopted:

- Treat histogram bins from HOG as a single "feature." The HOG feature types form nine features to a cell from histogram bins. Treating the nine features from HOG histograms as a single "feature" permits a single feature type to be removed from a cell on each iteration of backward selection.

- The percent correct classification performance metric from the analysis of the cell method introduced in Section 3.4 is used to evaluate the impact of removing a feature. The feature whose removal results in the highest percent correct classification is permanently removed from the feature set.

- Use the RVM classifier to evaluate the change in classification performance for the removal of a feature because the RVM classifier has a superior classification performance over LDA, as reported in Chapter III.

4.2.2 Method.

The same process as the analysis of the cell method introduced in Section 3.4 is used to evaluate the classification performance associated with removing different features. All the permutations of the feature set, denoted as $\mathbf{FS}_{ii}$, with a single feature removed are the feature vectors, denoted as $\mathbf{FV}_{\mathbf{v}}$, $\mathbf{v} \in [1...N - ii + 1]$. The classifier is trained and tested T times for each of the feature vectors, $\mathbf{FV}_{\mathbf{v}}$. The mean of the classification performance across all T trials for a feature vector, $\mathbf{FV}_v$, is the classification performance, Pc_v, of the feature vector.

A wrapper is added to the cell method introduced in Section 3.4 to systematically remove features one at a time. To systematically reduce the dimension of the feature set, $\mathbf{FS}_{ii}$, the wrapper permanently removes one feature from the feature set on each iteration and redefines all permutations, $\mathbf{FV}_{\mathbf{v}}$, of the feature set, $\mathbf{FS}_{ii+1}$. The process shown in Figure 4.1 summarizes the process used to execute backward selection.

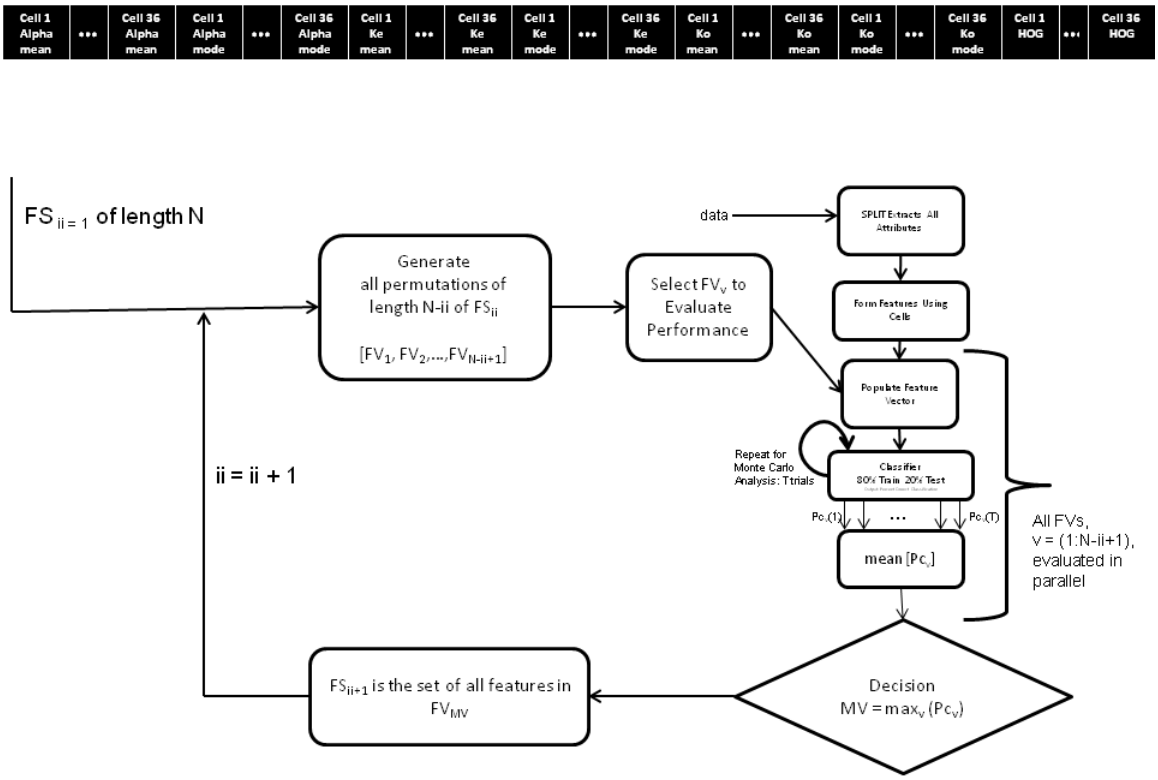

Figure 4.1: Block Diagram of the Backward Selection Process. Backward selection uses the same process as cell method to evaluate the classification performance of feature vectors.

The computational limits are observed by using fewer states than the analysis of the cell method implemented in Section 3.3. The states are sampled in elevation every three degrees, as opposed to the 0.25 degrees used in analysis of the cell method. By reducing the sampling in elevation of the states, the classification performance from the cell method decreases to the performance reported in Table 4.1.

Backward selection is initialized using the set of features in feature vector 20 of Table 4.1. The feature types in feature vector 20 are α mean, α mode, k_e mean, k_e mode, k_o mean, k_o mode, and HOG of amplitude. We do not initialize with feature vector 21 because the inclusion of the HOG of k_e and the HOG of k_o feature types increases the number of features

Table 4.1: Feature Vector Classification Results Using RVM With HOG Normalization, 3 Degree Elevation Sampling.

Feature Vector	α mean	α mode	K_e mean	K_e mode	K_o mean	K_o mode	amp HOG	k_e HOG	k_o HOG	Performance % Correct
1	√									60.45
2			√							72.53
3					√					76.16
4		√								58.04
5				√						66.27
6						√				73.30
7							√			75.35
8								√		65.28
9									√	68.45
10			√		√					77.83
11				√		√				76.89
12	√		√		√					66.41
13		√		√		√				74.41
14	√						√			74.58
15		√					√			73.45
16			√		√		√			80.40
17				√		√	√			82.12
18	√		√		√		√			77.65
19		√		√		√	√			80.47
20	√	√	√	√	√	√	√			77.23
21	√	√	√	√	√	√	√	√	√	79.89

by 1800 features. The additional 1800 features almost double the number of permutations of $\mathbf{FS}_{ii}$ to be evaluated. Doubling the number of permutations, doubles the computation

resources required for each iteration of backward selection. Also, because feature vector 17 has a higher corresponding classification performance, we know there is a more salient subset of features within feature vector 20.

4.2.3 Metrics.

The comparison of feature type saliency is based on the change in classification performance from the removal of the features corresponding to each feature type. Saliency of a feature type is expressed as

$$S_{FT} = \frac{\Delta CP_{FT}}{F_{FT}}, \tag{4.1}$$

where ΔCP_{FT} is the overall change in percent correct classification from the removal of the features of the feature type, and F_{FT} is the number of features removed of the feature type.

The comparison of cell saliency is done in a similar fashion as the saliency of feature types. The change in percent correct classification from the removal of the features corresponding to each cell is the metric used. Saliency of a cell is expressed as

$$S_C = \frac{\Delta CP_C}{C_C}, \tag{4.2}$$

where ΔCP_C is the overall change in classification from the removal of the features from the cell, and C_C is the number of features removed from the cell.

4.3 Results

Using the method from Section 4.2, we executed backward selection on the feature set from feature vector 20 reported in Table 4.1. The process of backward selection iterated 204 times to select a set of features that is more salient than the feature vectors reported in Table 4.1. The reported metrics indicate the least salient feature type is α mean, and the most salient feature type is k_o mode. The highest classification performance of each iteration of backward selection is shown in Figure 4.2.

Application of backward selection improves the classification performance, showing an increase from 77.23 to 84.28 percent correct classification. The first iteration achieved a

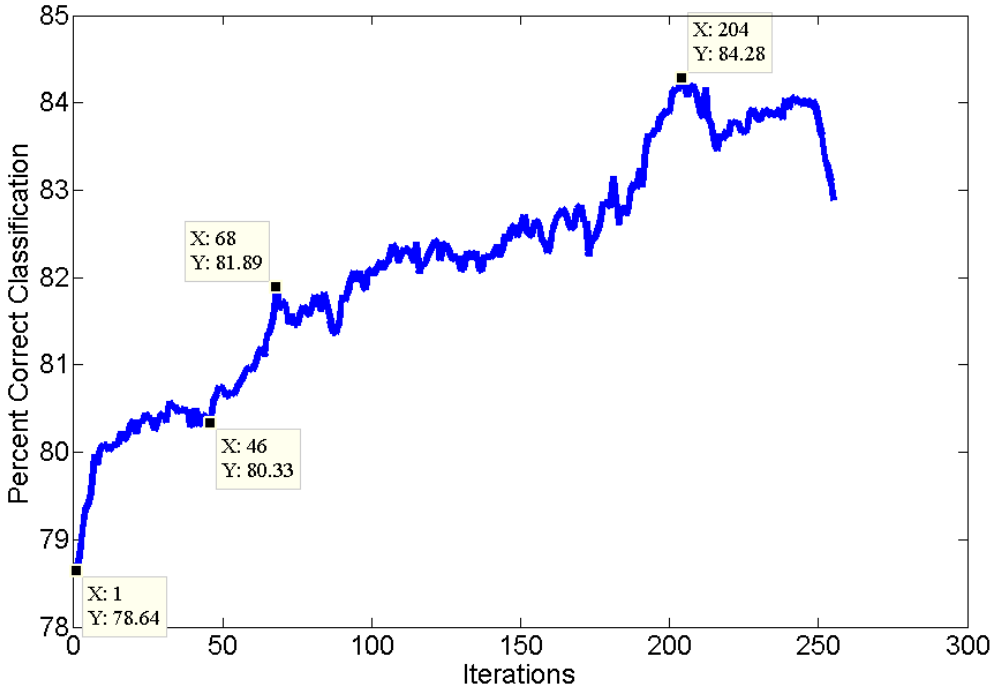

Figure 4.2: Tracking of the Highest Classification Performance of Each Iteration of Backward Selection Using RVM and 30 Trials.

classification performance of 78.64 percent correct classification with the removal of an α mode feature from cell 16. The first 20 iterations eliminated $\frac{6}{72}$ of the frequency response features, $\frac{5}{72}$ of the even bounce polarization response features, $\frac{5}{72}$ of the odd bounce polarization response features, and $\frac{4}{100}$ of the HOG feature sets with an increase of three percent correct classification. The second 20 iterations eliminated none of the frequency response features, $\frac{8}{67}$ of the remaining even bounce polarization response features, $\frac{8}{67}$ of the remaining odd bounce polarization response features, and $\frac{4}{100}$ of the remaining HOG features sets with no significant gain in percent correct classification. The next 40 iterations eliminated $\frac{6}{66}$ of the remaining frequency response features, $\frac{17}{59}$ of the remaining even bounce polarization features, $\frac{11}{59}$ of the remaining odd bounce polarization features, and $\frac{6}{92}$ of the remaining HOG feature sets with an increase of five percent correct classification

from the baseline. The next 40 iterations eliminated $\frac{2}{60}$ of the remaining frequency response features, $\frac{13}{42}$ of the remaining even bounce polarization features, $\frac{12}{48}$ of the remaining odd bounce polarization features, and $\frac{13}{86}$ of the remaining HOG feature sets with an increase to 82.32 percent correct classification. The next 60 iterations eliminated $\frac{14}{58}$ of the remaining frequency response features, $\frac{14}{29}$ of the remaining even bounce polarization features, $\frac{13}{36}$ of the remaining odd bounce polarization features, and $\frac{19}{73}$ of the remaining HOG feature sets with no significant increase in percent correct classification. The next 24 iterations eliminated $\frac{14}{44}$ of the remaining frequency response features, $\frac{4}{15}$ of the remaining even bounce polarization features, $\frac{4}{23}$ of the remaining odd bounce polarization features, and $\frac{2}{54}$ of the remaining HOG feature sets with an increase to the peak performance of 84.28 percent correct classification. From the analysis of the removed features and the change in classification performance, the saliency metric of feature types is shown in Figure 4.3.

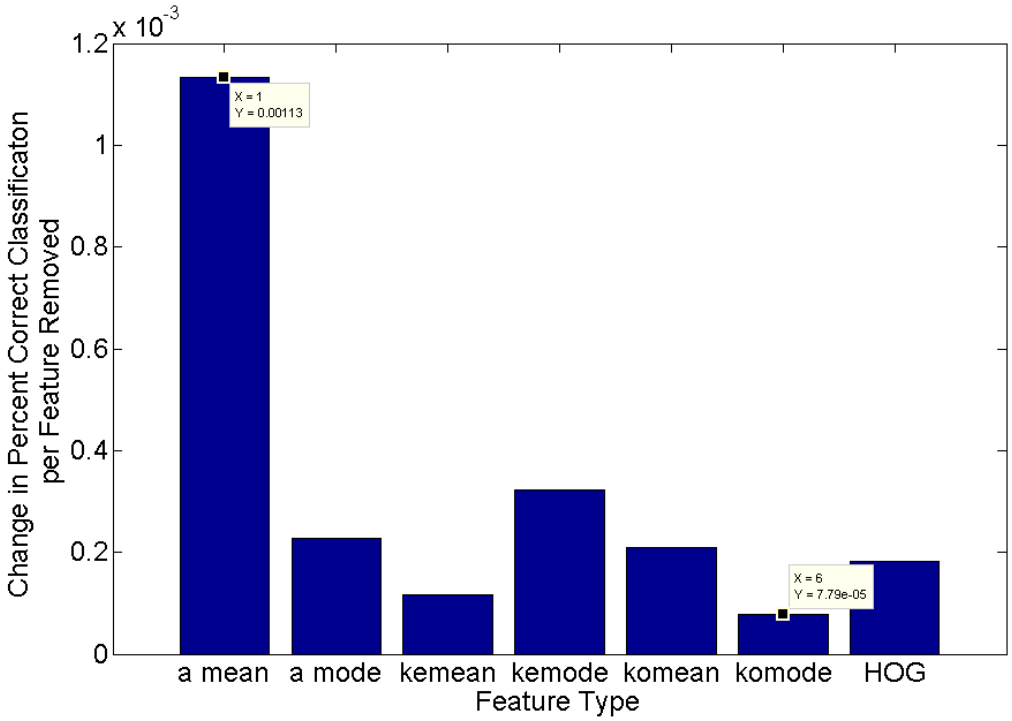

Figure 4.3: Comparison of Feature Type Saliency.

The least salient feature type is α mean. The greatest improvement in classification performance was made through the removal of the α mean features. Removal of any other feature type results in an increase in classification performance of less than 0.00032 percent correct classification on average. The increase in classification performance associated with the removal of the α mean features indicates that the feature had a negative impact on classification performance. The lack of saliency associated with α mean is consistent with the results from Chapter III.

The most salient feature type is k_o mode. The smallest improvement in classification performance was made through the removal of k_o mode features. Removal of the k_o mode type of features resulted in an increase of 7.79×10^{-5} percent correct classification on average. The k_o mode feature type's saliency is consistent with the results from Chapter III.

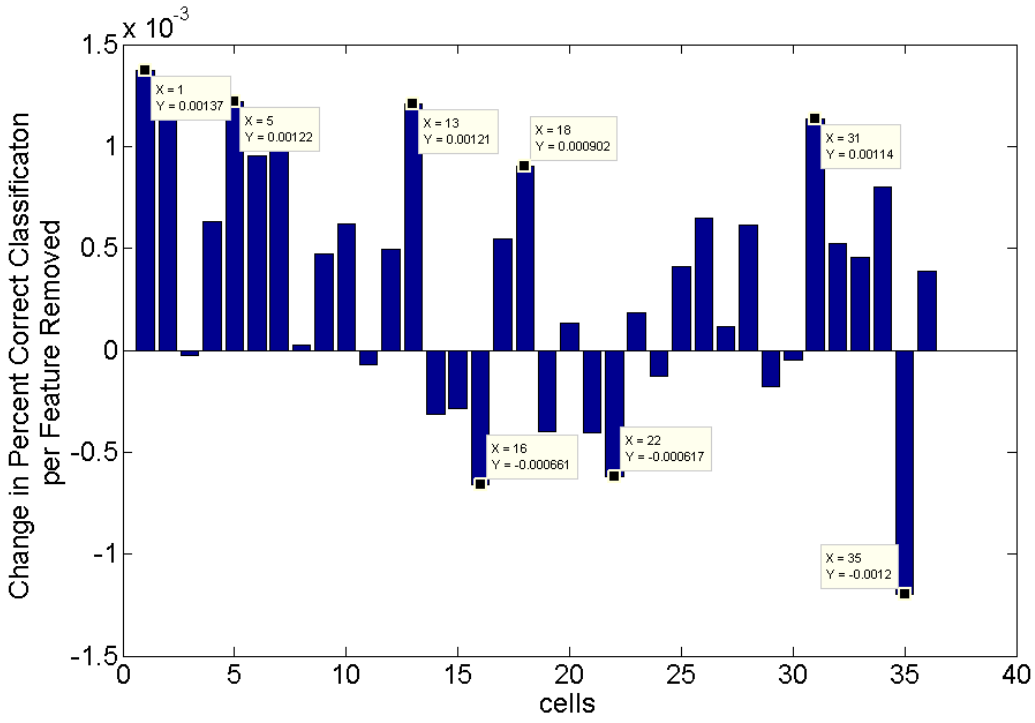

Figure 4.4: Comparison of Cell Saliency.

The saliency of cells is reported using a similar metric as the saliency of feature types. The saliency of cells is shown in Figure 4.4. The least salient cells are 1, 2, 5, 6, 7, 13, 18, and 31. Removal of features from cells 1, 2, 5, 6, 7, 13, 18, and 31 resulted in an increase of more than 0.00090 percent correct classification on average. The increase in classification performance associated with the removal of features is an indication that the features have a negative impact on classification performance. Removal of all other cells' features resulted in less improvement to classification performance than the removal of features from cell 18.

The most salient cells are 16, 22, and 35. Removal of features from cells 16, 22, and 35 resulted in a decrease of more than 0.00062 percent correct classification on average. The overall decrease in classification performance associated with the removal of features is an indication that the features have a positive impact on classification performance. Results show saliency for some cells.

Due to the symmetry of target vehicles, the saliency of the cells should also be symmetric. However, results indicate that the cell saliency is not symmetric. The lack of symmetry of the cells is an indication that the removal of non-salient features is hindered by the variance in classification performance. The variance in classification performance for the initial set of features is shown in Figure 4.5. Variance in classification performance at 30 trials is much greater than the maximum change observed in classification performance for each iteration. The maximum change in classification performance is 0.0045 percent correct classification. The hypothesis is that the ambiguity, resulting from variance, causes the wrong features to be removed and leads to the lack of symmetry in the saliency of the cells shown in Figure 4.4.

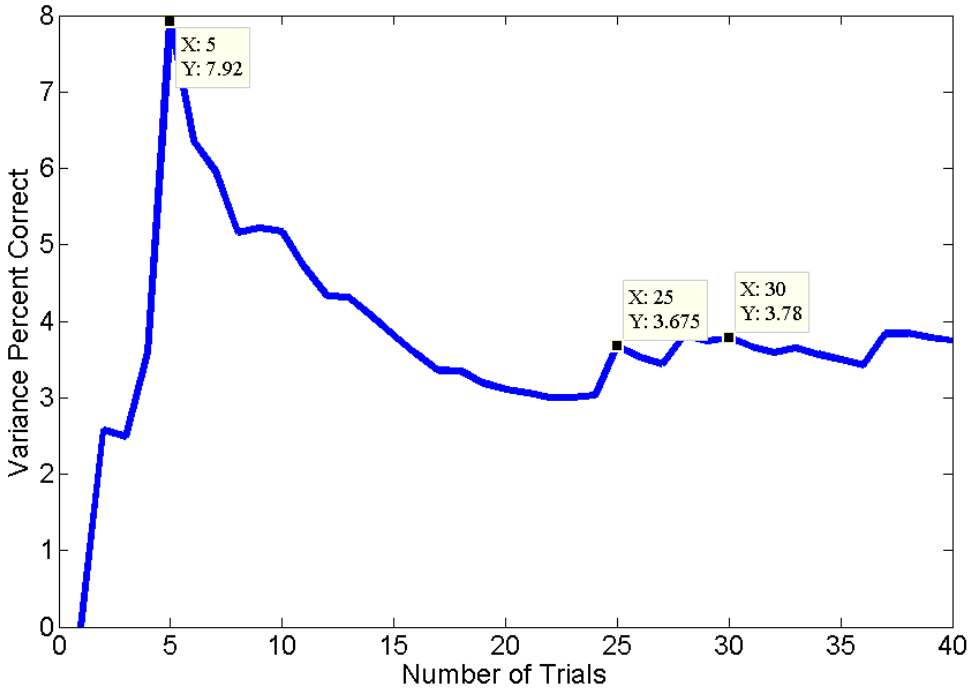

Figure 4.5: Variance in Classification Performance of Feature Vector 20.

V. Conclusions and Future Work

5.1 Conclusions

The cell method for feature extraction extracts features from SAR images related to the structure of physical elements of the target as well as captures image attributes from an entire image in a single feature vector. We implemented backward selection on the set of features formed with the cell method. Backward selection identified a set of features more salient than the initial set of features. The initial set had a classification performance of 77.23 percent correct classification, and the classification performance of the set of features selected by backward selection has a classification performance of 84.28 percent correct classification.

Backward selection selected a set of features that is not the most salient set of features. The set of features selected is more salient than the initial set in feature vector 20, but the improvement in classification performance was not monotonically increasing. We expect a monotonic increase in the classification performance with the removal of non-salient features [22]. Additionally, contrary to our expectations, the saliency of cells was not symmetric. We hypothesize that a reduction in classification variance will result in the selection of a more salient set of features. Completely eliminating the impact of the variance would result in the selection of the most salient set of features. High variance in the classification performance limits the performance of backward selection. Despite the high variance, the saliency of feature types is consistent with the results from the pixel method and the cell method, supporting the use of the saliency metric.

Analysis in Sections 3.2, 3.5, and 4.3 shows the frequency response attribute, α, is the least informative attribute for classifying SAR images from the CV Domes data set. Reported in the pixel method results, the inclusion of α in a feature vector decreased the corresponding classification performance of the feature vector. Reported in the cell method

results, the α feature type has the lowest corresponding classification performance of all the feature types. Finally, the saliency metric from the result of backward selection indicates the α feature type is the least salient feature type.

Analysis in Sections 3.5 and 4.3 shows that a combination of polarization response features and image amplitude features form the most salient set of features for classifying SAR images from the CV Domes data set. In the pixel method analysis, the inclusion of polarization and location features corresponded to the highest classification performance. In the cell method, the combination of polarization and HOG of amplitude feature types corresponded to the highest classification performance. The salient nature of amplitude features is expected as previous work on SAR ATR focuses on amplitude of images [5, 6]. The salient nature of polarization features, supports the incorporation of polarization information into SAR ATR.

5.2 Future Work

There are additional directions this research may take to follow what has been performed here. The directions span from a continuation of work, to the application of research. Future work may look into the variance in the classification performance, verify the extraction of the α attribute, review the SVM classifier for SAR ATR, and apply polarization attributes to other SAR ATR algorithms such as those in [3, 5, 6].

Future work should look into ways to manage or reduce the variance. We identified the variance in the classification performance as a driver for missing the most salient set of features in the execution of backward selection. One way to manage the variance is to use a greater number of trials in evaluating the classification performance. Also, using a greater number of target states may also reduce the variance. Both of these methods will require greater computational resources than used in this thesis.

The extraction of the α attribute should be investigated for accuracy. The poor performance of the α attribute may be attributed to either a bad extraction of the attribute

or a lack of consistency in the curvature of the physical elements of the vehicles. For future work to use the α attribute, the method for extracting the attribute should be verified.

Given the great amount of complexity in the SAR ATR problem, it may be advantageous to use the SVM classifier instead of the RVM classifier. SVM uses a greater number of support vectors than RVM. The greater number of support vectors allows for a hyperplane to mold to the high complexity of the feature space.

The polarization attributes, or a variation of them, should be applied to other SAR ATR algorithms. Results from the cell method, the pixel method, and backward selection identified the saliency of the polarization attributes in this thesis. The inclusion of polarization attributes may have a significant improvement in the performance of the algorithms.

Bibliography

[1] K. E. Dungan, C. Austin, J. Nehrbass, and L. C. Potter, "Civilian Vehicle Radar Data Domes," in *Algorithms for Synthetic Aperture Radar Imagery XVII, Proceedings of SPIE*, April 2010, vol. 7699, p. 76990P.

[2] T. D. Ross, S. W. Worrell, V. J. Velten, J. C. Mossing, and M. L. Bryant, "Standard SAR ATR evaluation experiments using the MSTAR public release data set," in *SPIE Algorithms for Synthetic Aperture Radar Imagry V*, September 1998, vol. 3370, pp. 556–573.

[3] C. Lin, B. Wang, X. Zhao, and M. Pang, "Optimizing kernel PCA using sparse representation-based classifier for MSTAR SAR image target recognition," *Mathematical Problems in Engineering*, vol. 2013, pp. 1–10, April 2013.

[4] A. K. Mishra and B. Mulgrew, "Radar Signal Classification Using PCA-Based Features," in *IEEE International Conference on Acoustics, Speech and Signal Processing* , May 2006, vol. 3, pp. III–1104 – III–1107.

[5] L. M. Novak, G. J. Owirka, and A. I. Weaver, "Automatic Target Recognition Using Enhanced Resolution SAR Data ," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 35, no. 1, pp. 157 – 175, 1999.

[6] K. E. Dungan and L. C. Potter, "Classifying vehicles in wide-angle radar using pyramid match hashing," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 3, pp. 577 – 591, 2011.

[7] M. S. Flynn, "Salient Feature Identification and Analysis Using Kernel-Based Classification Techniques for Synthetic Aperture Radar Automatic Target Recognition," M.S. thesis, Air Force Institute of Technology, 2014.

[8] C. V. Jakowatz, D. E. Wahl, P. H. Eichel, D. C. Ghiglia, and P. A. Thompson, *Spotlight-Mode Synthetic Aperture Radar: A Signal Processing Approach*, Kluwer Academic Publishers, Norwell, MA, 1996.

[9] Department of Defense High Performance Computing Modernization Program, "Spirit," http://centers.hpc.mil/systems/unclassified.html/Spirit, [Online; accessed February-2015].

[10] B. Bhanu, "Automatic target recognition: State of the art survey," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 22, no. 4, pp. 364–379, July 1986.

[11] D. F. Fuller, *Phase History Decomposition for Efficient Scatterer Classification in SAR Imagery*, Ph.D. thesis, Air Force Institute of Technology, 2011.

[12] V. C. Chen and H. Ling, *Time-Frequency Transforms for Radar Imaging and Signal Analysis*, Artech House, Inc, Norwood, Massachusetts, 2002.

[13] D. K. Saini, J. Smith, and M. A. Saville, "Vehicle Classification from Attributed SAR Imagery," *submitted to IET Radar, Sonar, and Navigation, Jan 2015*.

[14] E. F. Knott, J. F. Shaeffer, and M. T. Tuley, *Radar Cross Section*, Artech House, Inc, Norwood, Massachusetts, 1993.

[15] J. B. Keller, "Geometrical Theory of Diffraction," *J. Opt. Soc. Am.*, vol. 52, no. 2, pp. 116–130, February 1962.

[16] M. J. Gerry, L. C. Potter, I. J. Gupta, and A. Van Der Merwe, "A parametric model for synthetic aperture radar measurements," *IEEE Transactions on Antennas and Propagation*, vol. 47, no. 7, pp. 1179–1188, July 1999.

[17] M. A. Saville, J. A. Jackson, and D. F. Fuller, "Rethinking vehicle classification with wide-angle polarimetric SAR," *IEEE Aerospace and Electronic Systems Magazine*, vol. 29, no. 1, pp. 41–49, January 2014.

[18] C. A. Balanis, *Advanced Engineering Electromagnetics*, John Wiley & Sons, Inc., Hoboken, NJ, 2012.

[19] H. Mott, *Remote Sensing with Polarimetric Radar*, John Wiley & Sons, Inc., Hoboken, NJ, 2007.

[20] M. A. Richards, J. A. Scheer, and W. A. Holm, *Principles of Modern Radar*, SciTech Publishing, Edison, NJ, 2010.

[21] E. Krogager, "Advances of techniques for utilizing polarimetric features of radar targets," *NATO RTO SET Symposium on Target ID and Recognition Using RF Systems*, vol. 080-40, pp. 1–8, 2004.

[22] E. Alpaydin, *Introduction to Machine Learning*, The MIT Press, Cambridge, Mass, 2004.

[23] M. E Tipping, "Sparse Bayesian learning and the relevance vector machine," *Journal of Machine Learning Research*, vol. 1, pp. 211–244, September 2001.

[24] T. Fletcher, "Support Vector Machines Explained," http://www.tristanfletcher.co.uk /SVM20Explained.pdf/, [Online; accessed October-2014].

[25] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines*, Cambridge Univeristy Press, Cambridge, UK, 2000.

[26] K. Muller, S. Mika, G. Ratsch, K. Tsuda, and B. Scholkopf, "An Introduction to Kernel-Based Learning Algorithms," *IEEE Transactions on Neural Networks*, vol. 12, no. 2, pp. 181–201, 2001.

[27] T. Hastie, R. Tibshirani, and J. Friedman, *The elements of Statistical Learning*, Springer, New York, 2001.

[28] T. Fletcher, "Relevance Vector Machines Explained," http://www.tristanfletcher.co.uk /RVM20Explained.pdf/, [Online; accessed October-2014].

[29] US Air Force, "E-8 Joint Stars," http://www.af.mil/AboutUs/FactSheets/Display/ tabid/224/Article/104507/e-8c-joint-stars.aspx, 2005, [Online; accessed 04-March-2015].

[30] R. L. Moses, L. C. Potter, and M. Cetin, "Wide-angle SAR imaging," in *SPIE Algorithms for Synthetic Aperture Radar Imagery XI*, September 2004, vol. 5427, pp. 164–175.

[31] D. E. Dudgeon, R. T. Lacoss, C. H. Lazoot, and J. G. Verly, "Use of persistent scatterers for model-based recognition," in *Algorithms for Synthetic Aperture Radar Imagery , Proceedings of SPIE*, April 1994, vol. 2230, pp. 356–368.

[32] C. Nevill-Manning, G. Holmes, and I. Witten, "The Development of Holte's 1R Classifier," http://www.cs.waikato.ac.nz/ml/publications/1995/Nevill-Manning95-1R.pdf/, [Online; accessed March-2015].

[33] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, June 2005, vol. 1, pp. 886–893.

[34] MATLAB, "Extract histograms of oriented gradients (HOG) features," http://www. mathworks.com/help/vision/ref/extracthogfeatures.html/, 2014, [Online; accessed 24-October-2014].

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704–0188

1. REPORT DATE *(DD–MM–YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From — To)*
26–03–2015	Master's Thesis	Oct 2014–Mar 2015

4. TITLE AND SUBTITLE

Analysis of Features for Synthetic Aperture Radar Target Classification

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

McCauley, Aaron K., First Lieutenant, USAF

5d. PROJECT NUMBER

LRIR12RY19COR

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Air Force Institute of Technology
Graduate School of Engineering and Management (AFIT/EN)
2950 Hobson Way
WPAFB, OH 45433-7765

8. PERFORMING ORGANIZATION REPORT NUMBER

AFIT-ENG-MS-15-M-032

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Air Force Office of Scientific Research
Dr. Michael Kendra
875 N. Randolph St., Suite 325
Arlington, VA 22203
(703) 588-0671
michael.kendra@afosr.af.mil

10. SPONSOR/MONITOR'S ACRONYM(S)

AFOSR/RTA

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

DISTRIBUTION STATEMENT A:
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Considering two classes of vehicles, we aim to identify the physical elements of the vehicles with the most impact on identifying the class of the vehicle in synthetic aperture radar (SAR) images. We classify vehicles using features, from polarimetric SAR images, corresponding to the structure of physical elements. We demonstrate a method which determines the most impactful features to classification by applying subset selection on the features. Determination of the most impactful elements of the vehicles is beneficial to the development of low observables, target models, and automatic target recognition (ATR) algorithms.

We show how previous work with features from individual pixels is applied to a greater number of target states. At a greater number of target states, the previous work has poor classification performance. Additionally, the nature of the features from pixels limits the identification of the most impactful elements of vehicles. We apply concepts from optical sensing to reduce the limitation on identification of physical elements.

We draw from optical sensing feature extraction with the use of Histogram of Oriented Gradients (HOG). From the cells of HOG, we form features from frequency and polarization attributes of SAR images. Using a subset set of features, we achieve a classification performance of 96.10 percent correct classification. Using the features from HOG and the cells, we identify the features with the most impact.

Using backward selection, a process for subset selection, we identify the features with the most impact to classification. The execution of backward selection removes the features which induce the most error in classification. We report features extracted from polarization attributes of SAR images have the most positive impact on classification performance.

15. SUBJECT TERMS

SAR, ATR, HOG, RVM, LDA, SPLIT, salient feature identification, feature vector, extraction, classification, backward selection, subset selection

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Dr. Julie A. Jackson(ENG)
U	U	U	UU	80	19b. TELEPHONE NUMBER *(include area code)* (937) 255-3636 x4678 Julie.Jackson@afit.edu